University of BRISTOL

The Role of Whole Genome Duplication in the Evolution of the Land Plant Bodyplan

James W. Clark

University of Bristol
School of Earth Sciences
Life Sciences Building
24 Tyndall Avenue
BS8 1TQ

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science.

Supervised by Professor Philip Donoghue, Professor Harald Schneider & Dr Mark Puttick

49833 words

Abstract

The evolution of the land plant body plan has shaped the evolution of terrestrial ecosystems, human economics and the Earth's biosphere. The body plan has arisen through a series of innovations or 'jumps' that have in turn facilitated a greater diversity of architecture, reproductive complexity and the ability to occupy increasingly inhospitable environments. The evolution of novelty through gene duplication is a hypothesis that was first developed in the animal kingdom, though the more recent discovery of multiple whole genome duplication (WGD) events throughout plant evolutionary history has sparked a goldrush to identify and characterise WGD events, and to relate them to macroevolutionary hypotheses. As it stands, plants represent the best opportunity to establish a natural system in which to determine the outcomes of WGD events across disparate lineages. However, a fundamental requirement to studying WGD in a phylogenetic context is to first establish on which branch it occurred. Secondly, an accurate estimate of the absolute timing of the event can aid in providing a geological context. Finally, an effort must be made to capture and quantify the macroevolutionary outcome and determine the relative contribution of WGD. Studies of WGD to date have taken a 'tip down' approach, focussing solely on extant taxa and ignoring the wealth of information presented in the fossil record. In this thesis, I aim to establish and progress methods for the identification, dating and characterisation of WGD events in a palaeontological context. I establish a timescale for several of the most ancient duplication events in the most species rich lineages and the lineages on which we are most economically dependent. I demonstrate a means of measuring phenotypic diversity (disparity) at the kingdom level and use this to determine the relationship between WGD and morphological evolution. Ultimately, I show that the best approach to studying WGD in land plants is a holistic one, considering phylogenetic, developmental and palaeontological evidence.

Author Declaration

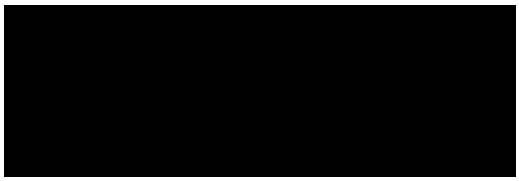
I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes, and that it has not been submitted for any other academic award. The work is entirely the candidate's, except where indicated in the text. All views expressed herein are the those of author.

James W. Clark

Statement of Collaboration

Chapters 2, 3 and 4 are research collaborations that have been led by the author. Experimental design, data collection, analyses, text and figure design were all carried out by the author. All remaining chapters are the work of the author, though in each case contributions of others are outlined at the beginning of each chapter.

Fossil calibrations are presented here, as in Chapters 2, 7 and Appendix 2, as they are integral to the analyses, though are in part based on the work of previous authors and only updated where necessary.



James W. Clark

Acknowledgements

I give thanks to my parents and my family, for offering constant refuge, warmth and grocery deliveries.

To Olivia, for the immeasurable love, support and foresight to look beyond a few ciders and a few more rats by the harbourside.

To Rebecca, for a seat in Royal Fort, a carrot and hummus sandwich and a brownie.

To Tom, Eliot, Amy and Leonie, for being especially quiet and respectful when walking up the stairs.

To James and Paul, for ensuring that my face is just about remembered in the Hillgrove.

To Frances, Holly and Alan, for validation that another round was indeed a good idea.

To Joe, Al and Ben, for being the wisest and most helpful of role models.

To Mark, for unlimited patience and companionship on distant shores.

To Phil, for the opportunity, motivation and guidance. I hope that I have done justice to your original designs.

I finally dedicate this thesis to the memory of William O. Langmaid and shall enjoy a Brandy & Benedictine when it's finally over.

Table of Contents

List of Figures

List of Tables

Thesis Outline

This thesis aims to elucidate the potential role of whole genome duplication (WGD) in the evolutionary history of land plants. The reasoning behind the focus on land plants is three-fold. First, the frequency of polyploidy and genome duplication in land plants is far greater than in animals, providing multiple natural experiments with which to test macroevolutionary hypotheses. Second, the evolution of land plants has been of crucial importance for the evolution of our planet. As architects of the environment, sources of food and other biotic interactions, understanding the evolution of plants is absolutely necessary to understand the evolution of life on land. The final reason is simply the author's own personal fascination with plants, and his belief that there is no finer way to spend four years than studying the patterns and causes of plant diversity.

The fossil record of land plants is far from perfect. Palaeobotanists and developmental biologists have gleaned impressive insight from the limited snapshots that we have of early plant evolution. However, the limitations of this fossil record should not prevent us from seeking to incorporate as much evidence as possible when studying plant macroevolution. Where possible I have sought to consider WGD in light of the fossil record, or else used the fossil record inform the timings of WGD.

The second chapter highlights shortcomings in the current state of research into WGD events. The deluge of sequence data is uncovering an ever-growing number of WGD events across the land plant phylogeny, where it is now more uncommon to report the absence of ancient polyploidy. I argue that these discoveries should be accompanied by appropriately constrained estimates of the phylogenetics and geological timing, and that such estimates can serve as the basis for hypotheses of macroevolution.

The third chapter is a study seeking to provide more accurate and precise estimates of the timing of the two most ancient WGD events observed in plants, shared by all seed plants and all angiosperms. I employ a phylogenomic approach to identify gene families with a signal of both duplication events and concatenate them into a single alignment. I also establish a set of fossil constraints across the plant phylogeny to calibrate a time tree. The results demonstrate that exploiting the signal present in multiple gene families can achieve estimates far more precise than previous attempts.

The fourth chapter displays the results of a collaborative effort to describe representatives of all major lineages of the plant kingdom using a large amount of discrete characters. These characters allowed me to calculate the dissimilarity between all taxa and represent these dissimilarities in multivariate space. The result is a two-dimensional representation of plant morphology across the kingdom. It reveals that plants have continued to push into novel regions of morphospace and that new bodyplans continue to evolve. By including fossil taxa, I show that some of the 'gaps' between lineages are the result of the extinction of intermediate forms, while others, and in particular the angiosperms, are truly distinct.

The fifth chapter applies the molecular clock dating methods described previously to three successive WGD events in one of the most successful and important lineages: the Poaceae. Taking an existing phylogenomic dataset, I confirm the phylogenetic position of each event, and point out an extensive history of WGD across multiple lineages within Poales. I show that none of the identified WGD events correlate with increased species diversification unless we accept an arbitrary lag phase, and that the rise to dominance of the grasses only occurred 40-50 million years after the genome duplication event.

The relationship between WGD and morphological evolution is difficult to study due to a lack of a quantitative framework to measure 'innovation'. In the sixth chapter I attempt to do so by using the morphological dataset established in Chapter 3 and the timing of genome duplication from Chapter 2. Evolutionary developmental studies have identified that the flower, the unique reproductive structure of angiosperms, may originate from a WGD event. I showed that while the flower, and by proxy WGD, has contributed significantly to the disparity of angiosperms, they remain a highly distinctive and diverse lineage even without. A survey of other WGD events and morphological datasets within the angiosperms shows a murky relationship between the two.

The seventh and final research chapter combines many of the techniques described previously, combining palaeontological, phylogenomic and transcriptomic approaches to identify, date and quantify the outcome of a truly ancient WGD event in Equisetaceae. This chapter highlights the need for a palaeontological approach, since it is only in the light of the fossil record that I show that the extant diversity which underwent the WGD event is only a small fraction of what was once a species rich and disparate lineage.

Chapter 1

Introduction

1 Summary

The evolution of land plants has in turn shaped the evolution of the Earth and the terrestrial biosphere. Recent studies have refined our hypotheses of the interrelationships among land plants and allowed a phylogenetic framework with which to study the processes that have shaped plant biodiversity. Plant genomes are incredibly diverse, and a phylogenetic survey shows that different lineages have evolved along contrasting trajectories. Polyploidy (whole genome duplication; WGD) describes any multiplication of the monoploid genome and is believed to be a driver of plant genomic, morphological and species diversity. It is rare among certain lineages, such as vertebrates, yet as much as 30% of some plant lineages are believed to be polyploids. The prevalence of polyploidy among land plants has long been viewed as a paradox. On one hand, polyploids appear prone to extinction – 'evolutionary dead ends' that form and rapidly disappear. On the other, it has recently been shown that nearly all lineages of extant land plants are descended from ancient polyploids, and that polyploid tends to precede the evolution of several major plant lineages, including all seed plants, angiosperms, eudicots and cereal crops. Phylogenomic approaches have revolutionised our ability to detect WGD events deep in the land plant phylogeny such that we are no longer asking how prevalent is polyploidy, but rather what is its significance? Answering this question requires more than simply placing each event on the phylogeny as we begin to introduce a comparative approach to studying the macroevolutionary consequences of polyploidy.

This chapter is unpublished.

2 The Evolutionary History of Land Plants

The emergence of plants onto land was a major episode in the evolution of the Earth. In recent years our understanding of the relationships among extant land plants and the timescale over which they have evolved has greatly improved (Fig 1.1). Phylogenomic data revealed the closest relative of land plants (Embryophyta), currently believed to be Zygnematales, a morphologically simple lineage of streptophyte algae (Wickett *et al.* 2014; Puttick *et al.* 2018). The evolution of streptophyte algae is now believed to be complex, with many lineages displaying a mosaic of traits that may have facilitated the eventual move to land (de Vries and Archibald 2018). Embryophyta originated 514.8-473.5 Ma and the interrelationships among lineages have recently been subject to revision, with a thorough consideration of phylogenomic data indicating that the non-vascular plants may be monophyletic (Puttick *et al.* 2018). Evidence for monophyly of all three non-vascular lineages was only weakly supported, yet strong support was recovered for a monophyletic lineage consisting of mosses and liverworts (Setaphyta). Support for Setaphyta crucially rejects the hypothesis that the liverworts were the earliest diverging lineage of land plants (Rensing 2018).

The vascular plant lineage (Tracheophyta) diverged from non-vascular relatives 450.8-430.4 Ma (Morris *et al.* 2018). Interrelationships among vascular plants are relatively stable, with a major division between Euphyllophyta (ferns and seed plants) and Lycophyta (clubmosses, spikemosses and quillworts) arising during 437.6-402.2 Ma (Morris *et al.* 2018). Within Euphyllophyta, the ferns are sister to the seed plants (Spermatophyta). The relationships among major lineages of ferns have also been revised, and the current hypothesis states that horsetails (*Equisetum*) are sister to a clade containing all other ferns (Shen *et al.* 2018).

Seed plants originated 365.0-329.8 Ma and consist two major lineages: the gymnosperms and angiosperms. Though some morphological data has supported a paraphyletic gymnosperms, phylogenomic data and more recent morphological

analyses support both groups as being monophyletic (Coiro *et al.* 2018; Ran *et al.* 2018). The most comprehensive phylogenomic analysis of gymnosperms supports a sister relationship between *Ginkgo* + Cycads which are in turn sister to all remaining gymnosperms (Ran *et al.* 2018). Gnetales are among the fastest evolving gymnosperms and have been variously resolved as sister to different gymnosperms and even angiosperms, yet are most recently placed as sister to Pinales (Ran *et al.* 2018).

The phylogeny of angiosperms is well resolved at the family level (APG IV). *Amborella,* a monotypic species restricted to the cloud forests of New Caledonia, is the basal-most lineage. Water lilies (Nympheales) and Austrobaileyales are the next most basal, together with *Amborella* forming the ANA grade. Remaining angiosperms form the Mesangiospermae, a highly diverse lineage. The major lineages within the Mesangiospermae diversified within a relatively short space of time, and consequently their interrelationships have been difficult to determine. A current hypothesis places the eudicots + *Ceratophyllum* as sister to the monocots, in turn sister to a Magnoliales + Chloranthales clade (APG IV).

3 The Diversity of Plant Genomes

Across the kingdom, plants harbour astonishingly diverse genomes. The genomes of different plant lineages appear to have evolved along very different trajectories, in terms of both genome size and chromosome number (Clark *et al.* 2016; Pellicer *et al.* 2018). Genome size (1C-value) varies over 2400-fold across land plants (Pellicer *et al.* 2018) and to date, the two largest eukaryotic genomes measured belong to two species of plant (Pellicer *et al.* 2010; Hidalgo *et al.* 2017). Non-vascular plants and lycophytes tend towards smaller genomes and with less variation (Pellicer *et al.* 2018). Ferns possess high mean genome sizes but are highly variable (Clark *et al.* 2016). Gymnosperms have the highest mean genome size alongside more conservative genome evolution (De La Torre *et al.* 2017). The genomes of angiosperms are highly dynamic and show the fastest rates of genome size evolution (Puttick *et al.* 2015). The

independent evolution of giant genomes across multiple plant lineages is likely a result of combinations of polyploidy, the expansion of repetitive elements and the conservation of DNA (Kelly *et al.* 2015).

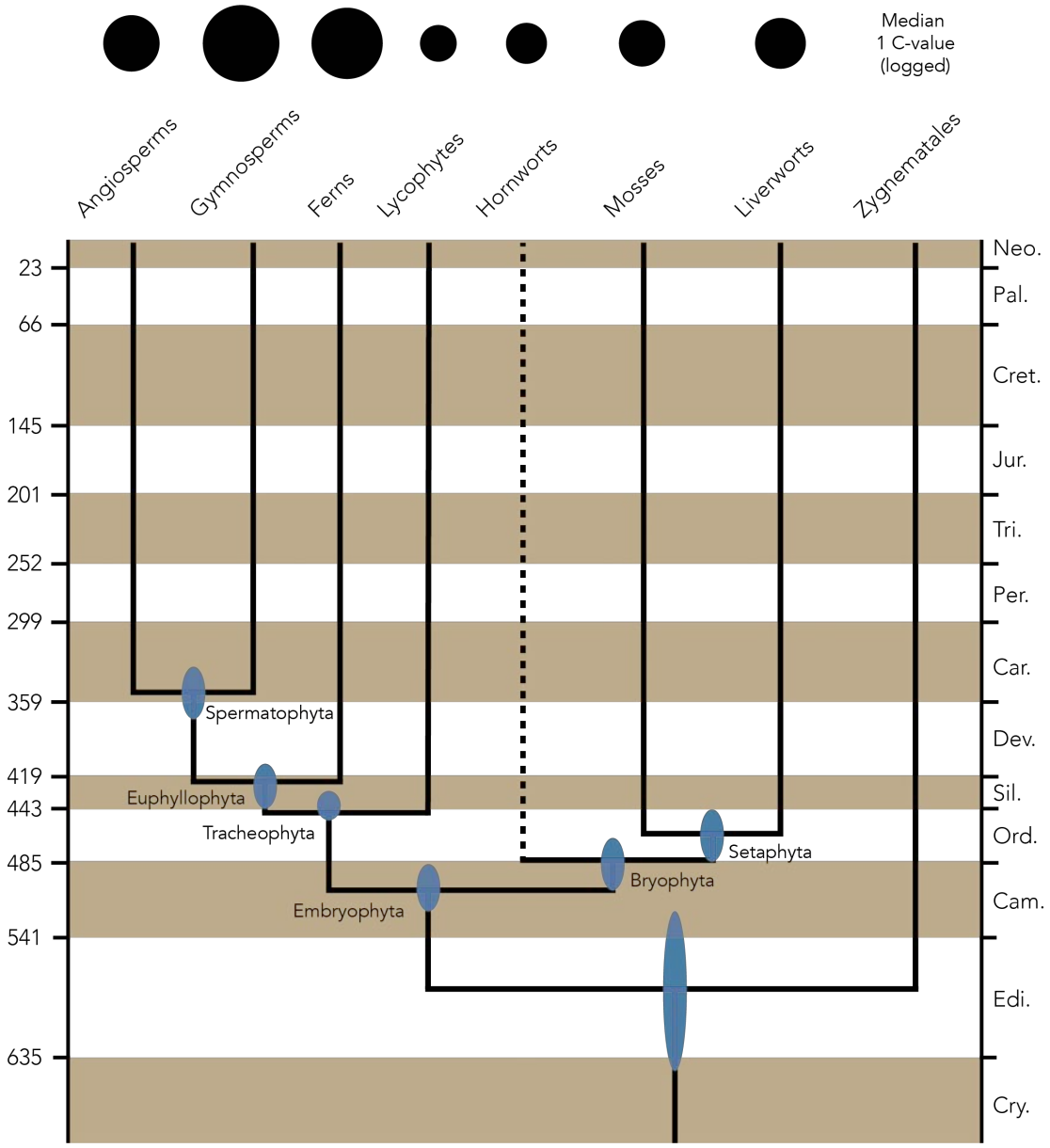


Figure 1.1 The evolutionary history of land plants (Embryophyta) reflecting current phylogenetic and divergence time hypotheses (Morris *et al.* 2018; Puttick *et al.* 2018). Uncertainty in the divergence time is represented by the 95% HPD (blue bars). The logged median genome size (megabases) is shown as a proportional circle above each lineage (Pellicer *et al.* 2018).

Trends in chromosome number evolution tend to reflect those of genome size: there is less variation in the non-vascular plants, conservatism in gymnosperms and fast rates of evolution in angiosperms and ferns (Clark & Puttick, unpublished). Interestingly, chromosome number and genome size are uncorrelated within all major lineages of land plants except ferns (Clark *et al.* 2016). This suggests some mechanism whereby ferns retain DNA after polyploidy and do not undergo the same process of diploidization as other plants, despite retaining a diploid expression pattern (Nakazato *et al.* 2006).

Such a brief survey of genome characteristics masks much of the variation that exists within these lineages, and indeed the mechanisms that have driven the diversity of plant genome architecture. Within the ferns alone multiple contrasting patterns of genome evolution are observed across clades, including gigantism, conservatism and high levels of variation (Clark *et al.* 2016). Within lineages there are always exceptions: the giant and slow-evolving genomes of most gymnosperms are contrasted to the (relatively) small and fast evolving genomes of Gnetales (Wan *et al.* 2018). In all, the landscape of plant genomics is highly diverse both in pattern and in process and the sequencing of increasingly diverse lineages is likely to yield greater insight into the mechanisms of genome evolution.

4 A History of Genome Duplication

Evolutionary biologists and palaeontologists have observed an increase in complexity across multiple lineages (Ruiz-Trillo and Nedelcu 2015). A question that arises from this is how organisms can evolve novel structures, tissues and organs with a constrained repertoire of genes. The intricacy of the eukaryotic genome allows morphological complexity to arise through multiple means, yet the duplication of existing genes provides one clear mechanism to explain the origins of novelty (Lewis 1951; Taylor and Raes 2004). The duplication of genes provides a second, redundant copy which has the potential to evolve new functions. The duplication of an entire

genome (whole genome duplication; WGD) provides a redundant copy of every gene, and in theory the opportunity for even greater novelty to evolve (Ohno 1970). The simultaneous duplication of each gene provided a more efficient means for the duplicates to co-evolve, and so Ohno proposed that WGD events, rather than small-scale or tandem duplications, were more critical to evolution of complexity.

Polyploidy can arise through various means, including somatic doubling and the fertilisation of unreduced gametes (Ramsey and Schemske 2002). Modes of polyploidy have been defined based on its origins and on its timing. Autopolyploids are formed from a single parent while allopolyploids form from a hybridisation between species followed by doubling. Auto- and allopolyploidy exist on a spectrum with intermediate cases of 'segmental allopolyploids' arising from hybridisation between varyingly divergent genomes. Polyploidy is also categorised based on when it occurred during a lineages' evolution. Neopolyploidy refers to recent polyploidy, while palaeopolyploidy describes events that are more ancient.

The potential role for WGD in generating evolutionary novelty centred around the discovery of two putative rounds of genome duplication in the ancestors of all vertebrates (Ohno 1970; Kasahara 2007). While controversial, the current hypothesis predicts one round of duplication in the ancestor of all vertebrates and a second in the ancestor of all gnathostomes (jawed fish) (Van de Peer *et al.* 2010; Hafeez *et al.* 2016). The timing of these events and the coincidence of the evolution of the vertebrate skeleton and jaw, respectively, provided circumstantial evidence for a role for WGD in promoting morphological complexity. What united most studies was a 'tip-down' approach, considering evidence provided only by extant taxa, and unsurprisingly most found that each WGD event preceded a rapid burst of morphological evolution (Sidow 1996). A palaeontological approach revealed a more complex pattern: that many of the morphological innovations associated with WGD had appeared in a stepwise manner, preserved in the fossil record (Donoghue and Purnell 2005).

Further WGD events have been uncovered in the ancestors of all teleost and salmonid fish (Allendorf and Thorgaard 1984; Hoegg *et al.* 2004), yet rather than morphological complexity, these events have mostly been explored in relation to species diversification. The differential loss of one gene duplicate (paralog) between individuals has the potential to drive reproductive isolation (Lynch and Force 2000). Experimental studies in yeast have confirmed the potential for reproductive isolation to arise through reciprocal gene loss (Maclean and Greig 2011), yet the contribution of gene loss to speciation rates on a macroevolutionary scale is questionable (Muir and Hahn 2015).

Tests relating WGD to macroevolution have mostly been perfomed in vertebrates, where they are limited by the small number of events. For example, the link between WGD and diversification in teleost fish is currently equivocal. Teleost fish represent the majority of vertebrate diversity, yet the bulk of teleost diversity is contained in lineages that diversified long after the WGD event (Alfaro *et al.* 2009). Likewise, the WGD shared by salmonid fish is decoupled from high rates of diversification by 40-50 Ma, making the link between the two tenuous at best, unless a lag period is accepted between the WGD event and the macroevolutionary outcome. The phylogenetic placement of the ancient WGD events, 1R and 2R, is controversial, and have never been directly linked to increased species diversification. Comparative tests of specific hypotheses relating to WGD and macroevolution are thus difficult in vertebrates where, despite the large volumes of genomic data, the incidence of WGD is too low.

5 Genome Duplication in the Plant Kingdom

For a long time it has been known that WGD is far more prevalent in the plant kingdom than in animals (Stebbins Jr 1940). Polyploidy is a means of instantly generating reproductive isolation and in turn greater species diversity. It is estimated that 15% of extant angiosperms and 31% of extant ferns are polyploids based on

chromosome counts (Wood *et al.* 2009). While the role of polyploidy in shaping extant patterns of diversity is accepted, there has been significant disagreement about the long-term effects of polyploidy, which has been framed as the 'polyploidy paradox'. Analyses of rates of diversification among polyploid and diploid lineages found that polyploids tend to diversify more slowly and suffer extinction more frequently than their diploid relatives (Arrigo and Barker 2012; Mayrose *et al.* 2015). This has painted the view that polyploids are 'evolutionary dead ends' and that polyploidy is at best a short-term strategy.

However, the distribution of ancient (palaeo-) polyploid events shows that many, if not most, of the most successful lineages of land plants are defined by lineage-specific WGD events. For example, all living members of the seed plants, conifers, flowering plants, core eudicots and most monocots have undergone lineage-specific WGD events (Vanneste *et al.* 2014b; Li *et al.* 2015; Clark and Donoghue 2018). The initial rarity of these ancient WGD events suggested that while on the whole polyploids were unsuccessful, in a few lineages WGD facilitated greater success – these were the 'rarely successful polyploids' (Arrigo and Barker 2012). As more sequence data has become available it has become increasingly apparent that the initial rarity of WGD events in plant evolutionary history was an artefact of low sampling density. With each sequenced genome or transcriptome, it is now more uncommon *not* to uncover a new WGD event. For example, it was recently demonstrated that the members of a single order, Caryophyllales, may have undergone as many as 26 ancient WGD events (Walker *et al.* 2017; Smith *et al.* 2018).

While current thinking has swung towards a role for WGD in promoting evolutionary success, many lineages remain overlooked. While many lineages that have undergone WGD appear highly diverse, many do not. For example, even species poor lineages, such as the dogwoods, show evidence of an independent WGD event (Yu *et al.* 2017), and WGD events in the ancient sphenopsids *Equisetum* (Vanneste et

al. 2015) and in *Sphagnum* peat mosses (Devos *et al.* 2016) are more difficult to relate to any proposed macroevolutionary hypothesis.

This contrary picture is surely part of what makes WGD such a fascinating phenomenon. Ultimately, given the frequency of WGD events within the plant kingdom, and that many more likely to be uncovered, plants represent the single best system with which to determine any role for WGD in macroevolution.

Chapter 2

Whole genome duplication and plant macroevolution

James W. Clark and Philip C.J. Donoghue

1 Summary

Whole genome duplication (WGD) is characteristic of almost all major lineages of land plants. Unfortunately, the timing of WGD events is loosely constrained and hypotheses of evolutionary consequence are poorly formulated, making them difficult to test. Using examples from across the plant kingdom, we show that estimates of timing can be improved through the application of molecular clock methodology to multigene datasets. Further, we show that phenotypic change can be quantified in morphospaces and that relative phenotypic disparity can be compared in the light of WGD. Together, these approaches facilitate tests of hypotheses on the role of WGD in plant evolution, effecting the potential of plants as a model system for investigating the role WGD in macroevolution.

This chapter is published as Clark, J.W. & Donoghue, P.C.J. 2018. 'Whole-Genome Duplication and Plant Macroevolution', *Trends in Plant Science.* 23: 933-945. *The article was conceived by J. Clark and P. Donoghue. The writing of the article and analyses therein were conducted by J. Clark and were further revised by P. Donoghue. The 'Box' elements have been incorporated into the main text.*

2 A history of Whole Genome Duplication in Land Plants

Whole genome duplication (WGD) encompasses multiple processes that lead to the formation of a polyploid organism with three or more sets of the base chromosome number. It has been invoked as a cause of macroevolutionary change (Ohno 1970), explaining everything from extinction resistance to fundamental evolutionary innovation. WGD has been proposed as a driver of diversity (Tank *et al.* 2015; Ren *et al.* 2018), herbivore interactions (Edger *et al.* 2015), geographic expansions (Barker *et al.* 2016), climatic niche shifts (Smith *et al.* 2018) and facilitating lineage longevity (Vanneste *et al.* 2015). Clustering of WGD events along the K-Pg boundary has led to the hypothesis that genome duplication may have facilitated evolutionary success in the wake of the end Cretaceous mass extinction event (Fawcett *et al.* 2009; Lohaus and Van de Peer 2016). Equally though, it is possible that the extensive history of WGD in plant evolution is incidental or inconsequential, and there are examples, such as mosses and horsetails (Vanneste *et al.* 2015; Devos *et al.* 2016), where a macroevolutionary scale phenotypic impact is not evident. Ancient WGD events (palaeopolyploidy) first appeared rare (Cui *et al.* 2006), yet newly sequenced genomes have revealed duplication in an increasing diversity of plant lineages (Walker *et al.* 2017; Xiang *et al.* 2017). However, with few exceptions, it appears that most of the hypothesised macroevolutionary outcomes have neither been tested nor formulated as hypotheses that are readily testable, despite the diversity of comparative methods available for facilitating such tests. There are multiple emerging models explaining how complexity and novelty may arise through genome duplication (Conant *et al.* 2014), although fundamental questions remain as to why the outcomes of WGD are so disparate among lineages and whether the nature of the ploidy event influences the outcome. Tests are needed to quantify the macroevolutionary change in the wake of WGD, or else we risk WGD becoming a phenomenon that explains everything and, therefore, nothing.

WGD has occurred across the breadth of eukaryote phylogeny (Donoghue and Purnell 2005; Marcet-Houben and Gabaldon 2015; Schwager *et al.* 2017; Li *et al.* 2018b), but the majority of WGD events have occurred within land plants (Embryophyta; Fig 2.1). As such, plants provide very many natural experiments from which it may be possible to develop a general theory on the role of WGD in macroevolution. Patterns of diversification among extant taxa have pointed towards a scenario of rarely successful polyploids (Arrigo and Barker 2012; Soltis *et al.* 2014b). However, all members of the most diverse lineage of land plants, the seed plants (Spermatophyta), are descended from an ancestor that underwent at least one round of WGD (Jiao *et al.* 2011; Li *et al.* 2015). Furthermore, within Spermatophyta, another WGD is shared by all flowering plants (angiosperms) (Jiao *et al.* 2011), as well as others shared in turn by several major clades of flowering plants including the monocots (Jiao *et al.* 2014), eudicots (Jaillon *et al.* 2007; Jiao *et al.* 2012), Asteraceae (Barker *et al.* 2016; Huang *et al.* 2016), Brassicales (Kagale *et al.* 2014), legumes (Cannon *et al.* 2015) and in the most economically important plants, the grasses (Estep *et al.* 2014; McKain *et al.* 2016) (Fig 2.2). The paucity of ancient WGD events that was perceived early in the history of genome sequencing is looking increasingly like an oversight, with denser sampling revealing multiple WGD events during the evolution of taxonomically large and small lineages (Smith *et al.* 2018).

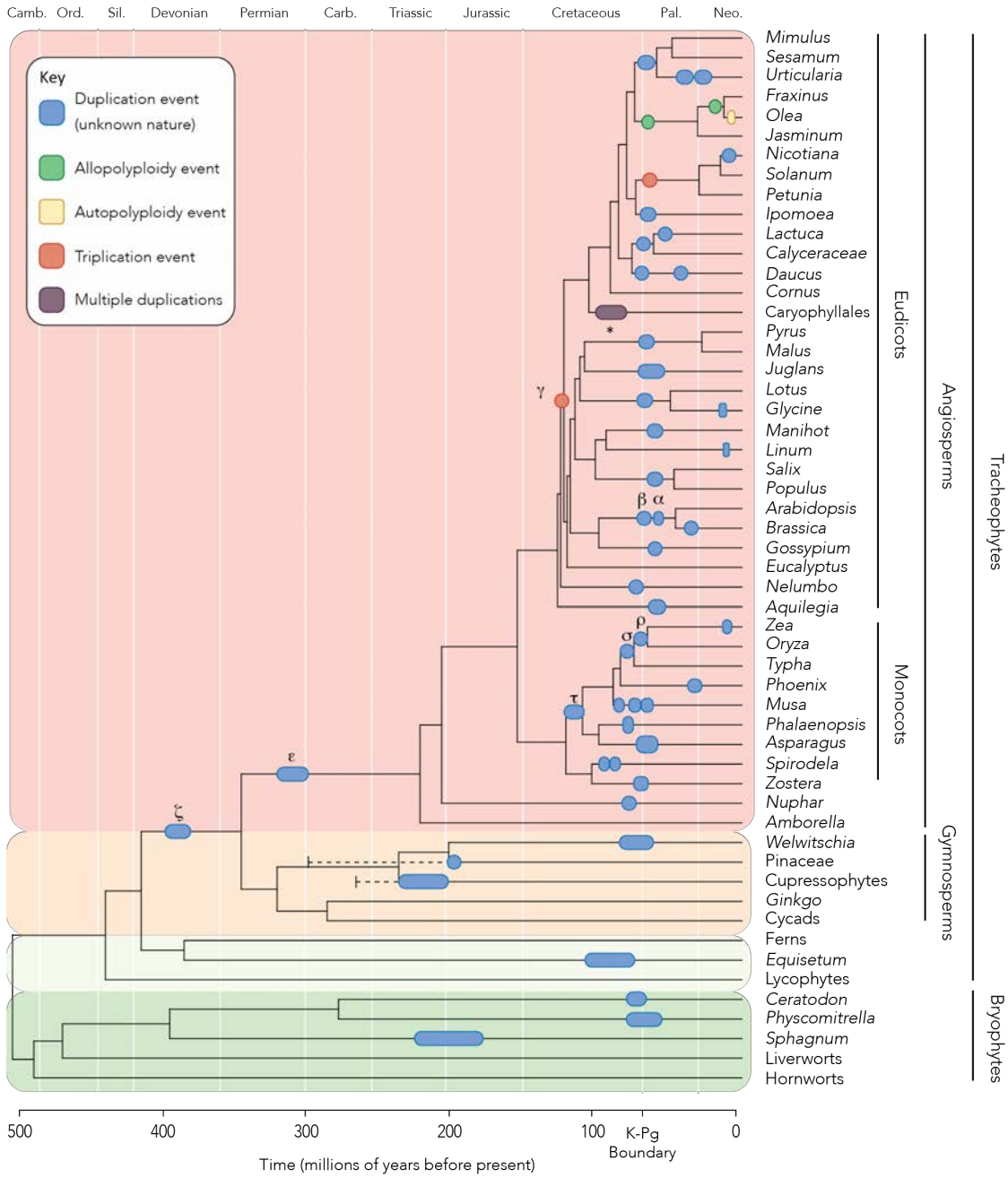


Figure 2.1. The distribution of known WGD events within the plant kingdom. Most events are shown from Van de Peer *et al.* (2017) but have been updated. The length of each bar along the branch indicates the current estimate for its age. Duplication events of unknown origin are shown in navy blue, triplications in red, known autopolyploidy events in yellow and allopolyploidy events in green. The white bar associated with Caryophyllales represents 26 independent WGD events, some of which are autopolyploidy and some allopolyploidy. Named duplication events are shown alongside their Greek letter.

3 Double Dates – the absolute timing of WGD

Hypotheses on the role of WGD in plant macroevolution are contingent on the phylogenetic (relative) and geological (absolute) timing of each event. Methods to identify WGD events are many and varied: paralog substitution distributions (*Ks* plots) (Lynch and Conery 2000; Vanneste *et al.* 2013), phylogenomics ((Jiao *et al.* 2011), genome size, karyotype, gene copy number analyses (Clark *et al.* 2016; Tiley *et al.* 2016), and synteny (Lyons *et al.* 2008; Tang *et al.* 2008; Jiao *et al.* 2014). Greater sampling of diversity helps resolve the phylogenetic (relative) timing of each WGD, yet to refine these hypotheses it is important that their absolute ages are estimated with accuracy and precision. Absolute ages can be constrained by bracketing the age of speciation events since WGD must have occurred after the divergence of species that have not undergone genome duplication and before those living species that have (Fig 2.2). When taxonomic sampling is dense and the WGD occurred on a short branch (such as with more recent events) this can yield relatively precise age estimates (Edger *et al.* 2018). However, with increasing uncertainty in species divergence time estimates, longer branches, monotypic lineages, or less dense sampling, it becomes more challenging to directly estimate the timing of a WGD.

As well as being a means to identify and relatively date WGD events, both *Ks* analyses and phylogenomic methods can be used to directly infer the age of WGD events (Lynch and Conery 2000; Blanc and Wolfe 2004; Vanneste *et al.* 2013; Vanneste *et al.* 2014b). *Ks* plots represent distributions of rates of synonymous substitutions between paralogs. A peak in the distribution is interpreted as a WGD event and distributions compared between species can reveal shared duplication events. An external calibration can convert *Ks* rates into geological time, though this is often done by comparing the position of the peak in *Ks* rates to ages inferred from phylogenomic dating, for example a *Ks* value of 0.6 and 1.1 synonymous substitutions per site corresponds to an age of 50 - 70 million years (Vanneste *et al.* 2014b). These methods assume a strict rate of molecular evolution, and different rates produce

highly variable age estimates. The signature of increasingly ancient WGD events is eroded by sequence saturation and so the detection of more ancient events leads to inaccuracy (Vanneste *et al.* 2013). For example, a WGD event predicted in the early-diverging gymnosperm *Ginkgo biloba* was estimated between 500-700 Ma - predating estimates for the origin of land plants (Guan *et al.* 2016; Roodt *et al.* 2017; Morris *et al.* 2018).

Phylogenomic approaches exploit the signal of paralogy present in the history of gene families to directly estimate the age of the WGD event (Jiao *et al.* 2011). This requires the reconstruction of gene families across multiple species (also termed a phylome (Huerta-Cepas *et al.* 2014) and subjecting them to molecular clock analysis. Molecular clock methodology has typically been applied to dating species divergences but can also be used to date both speciation and duplication events within gene trees. Typically, molecular clock analyses have investigated each gene family in isolation, producing both a topological and temporal estimate of WGD. Molecular clock approaches to dating WGD have either been flawed by the underlying algorithm (Ruprecht *et al.* 2017), or when more powerful Bayesian uncorrelated methods have been used, by the limited sampling of taxa and appropriate fossil calibrations (Vanneste *et al.* 2014b). Furthermore, dating individual gene families does not make best use of information available since individual gene families have low statistical power, yielding imprecise, if not inaccurate, estimates of gene and (by inference) genome duplication.

The paralog sets derived from a WGD share the same age and can be combined in a concatenated alignment that is capable of producing far more precise results than any single gene family (Macqueen and Johnston 2014; Clark and Donoghue 2017). Precision of estimated dates are not the sole concern – accuracy is important too, and it is achieved using conservative palaeontological constraints on speciation events (Parham *et al.* 2012), alongside clock methods that can model both the uncertainty in the fossil evidence and the variation in rates of evolution between

genes (Warnock *et al.* 2015; Clark and Donoghue 2017). Figure 2.2 shows a schematic analysis of the genome duplication present in the ancestor of all grasses (*rho*). This event is evident in the genomes and phylomes of multiple extant grass species which, due to their economic value as food crops, have been well-sampled by sequencing projects (McKain *et al.* 2016).

As well as being able to inform on the coincidence of WGD with geological or biogeographic events, these approaches co-estimate the timing of duplication alongside the timing of speciation. This allows us to see how early or late WGD occurred relative to the crown (extant) clade and to directly estimate lag between the WGD event and any hypothesized macroevolutionary consequences (Clark and Donoghue 2017).

3.1 WGD and K-Pg

The distribution of WGD events both across the plant phylogeny and through time has revealed that in multiple independent lineages WGD events appear to cluster along the K-Pg boundary (Fig 2.1). This has led to two related hypotheses: that genome duplication may have conferred an 'extinction resistance' to certain lineages of plants, and that polyploid genomes may have allowed surviving lineages to rise to dominance in the wake of the mass extinction.

Polyploid plants are sometimes found towards the edge of species ranges and polyploid genomes facilitate rapid radiations and invasiveness (te Beest *et al.* 2012). Polyploid genomes also possess a 'mutational robustness' relative to diploids which may provide short term advantages which may have allowed them to survive and then thrive. An alternative hypothesis suggests that it is not WGD itself that facilitated extinction resistance, but the coincidence that many newly formed polyploids rely on selfing to reproduce. Selfing is also associated with extreme or novel habitats, but in the long term is seen as an evolutionary dead end. A return to outbreeding could

allow the continued success of these lineages and may also explain the apparent lag between WGD and diversification.

These hypotheses are entirely dependent on the precise timing of each duplication event. As shown in Figure 2.2, current estimates for the timing of WGD is likely to change given a careful appraisal of the fossil record. As such, until each WGD event that lies close to the boundary is considered, this correlation should be treated with caution.

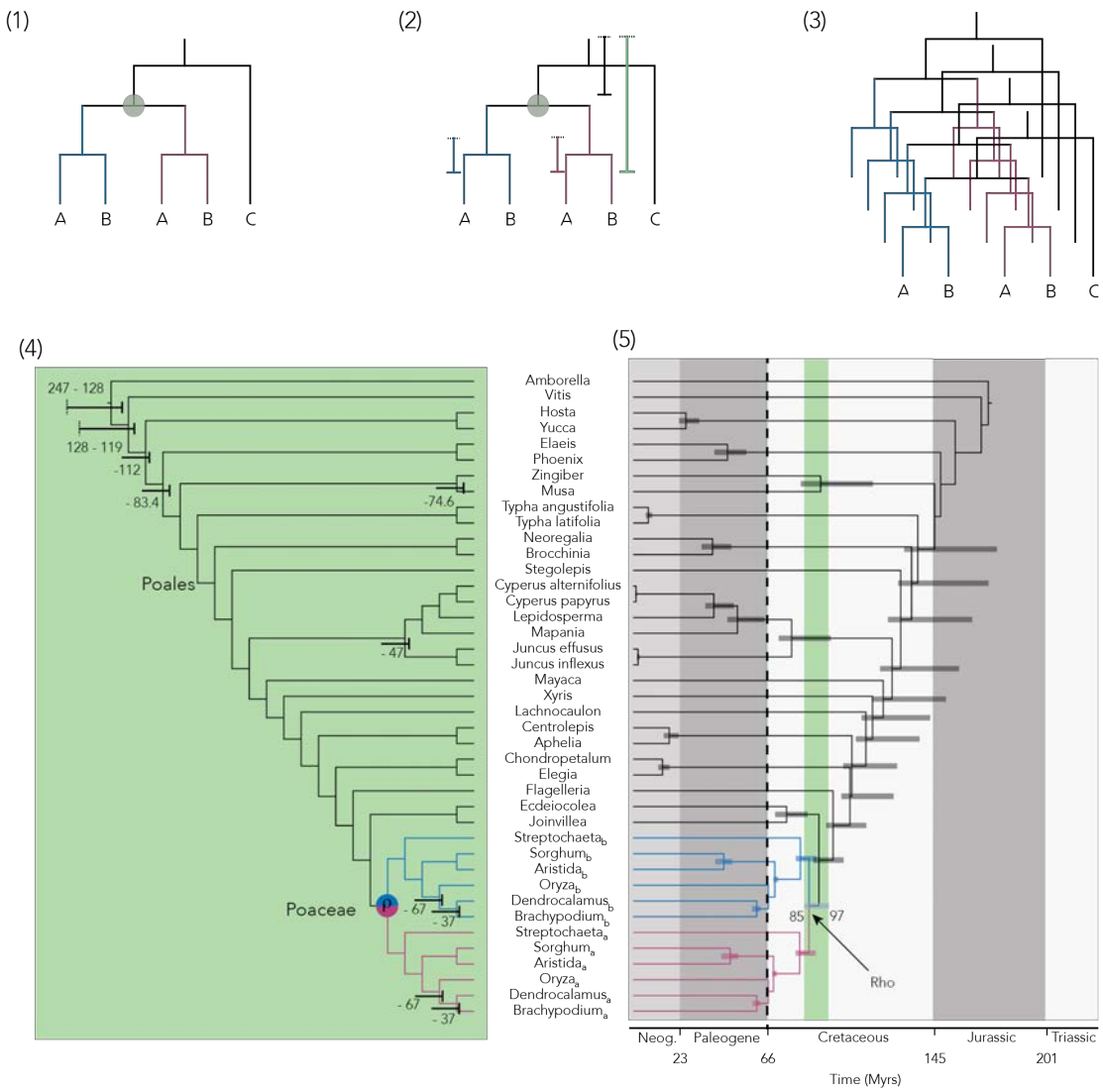


Figure 2.2. Dating WGD by combining genomics and the fossil record. 1) the history of WGD is present in individual gene families. Taxa A and B have undergone a shared duplication event, which taxon C has not. 2) The timing of the duplication is bracketed by the timing of the divergence of A and B and the divergence of A+B and C. These divergence times can be calibrated using distributions between minimum and soft maximum ages. 3) Multiple gene families with a shared signal of WGD can be concatenated to maximise the precision of the analysis. 4) Accuracy is achieved through a careful appraisal of the fossil record and by modelling uncertainty through soft maximum ages (Iles *et al.* 2015; Clark and Donoghue 2017). 5) A Bayesian molecular clock analysis reveals that the grass duplication (*rho*) occurred 85-97 Ma (95% HPD).

3.3 Dating whole genome duplication in grasses

Syntenic and phylogenomic evidence points towards a WGD event in the ancestor of all extant grasses (Poaceae(Jiao *et al.* 2014; McKain *et al.* 2016). The *rho* event has previously been dated through phylogenetic bracketing to ~ 70 Ma (Paterson *et al.* 2004) and is one of the numerous plant WGDs hypothesised to approximate the K-Pg boundary (Van de Peer *et al.* 2017). We sampled the gene families previously shown to retain the signal of the *rho* duplication and concatenated them into an alignment (Fig 2.2). Fossil evidence constrains the minimum age on speciation nodes, and in some cases can be used to apply 'soft' maxima (Donoghue and Benton 2007) (Fig 2.2). The Late Cretaceous fossil phytolith taxon *Changii indicum* is assigned to the crown group (i.e. the living clade) of the Oryzeae tribe and provides a minimum age of 66 Ma based on radiometric dating (Prasad *et al.* 2011; Christin *et al.* 2014; Iles *et al.* 2015). This fossil placement is contentious and can instead be used to calibrate the BOP+PACMAD (Bambusoideae, Oryzeae, Pooideae + Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonoideae) clade of grasses (Christin *et al.* 2014). We applied further fossil constraints and, combined with the concatenated alignment, these calibrations inform a Bayesian molecular clock analysis performed on the fixed topology of McKain *et al.* (2016) in MCMCTREE (Yang 2007). The results predict that the WGD took place between 97 to 85 Ma, and in this case is not compatible with the hypothesis that this event coincides with the K-Pg boundary (Fig 2.2).

4 Whole Genomes and Diversification

Diversification is one of the most widely proposed consequences of WGD in plants. This relationship has been explored at multiple levels across angiosperms yet support for a correlation remains equivocal (Soltis *et al.* 2009; Estep *et al.* 2014; Tank *et al.* 2015; Kellogg 2016). There is little evidence supporting a direct shift in diversification immediately following WGD. Instead, there is some support for the

proposed 'WGD lag-time' model, wherein diversification follows WGD but only after a protracted period of geological time (Tank *et al.* 2015). The lag period has been measured either as a period of absolute time or as an arbitrary measure of time such as the number of nodes separating a WGD event and a subsequent shift in the rate of diversification. When the age of the duplication event and the subsequent speciation events are co-estimated, the absolute age and duration of the lag can be estimated directly (Clark and Donoghue 2017). Estimates for the timing of the angiosperm-specific genome duplication event imply that it occurred 65-35 Myr before the divergence of crown angiosperms (the living clade of flowering plants), closer to 70 Myr before the radiation of Mesangiospermae and over 100 Myr before a detectable angiosperm radiation in the fossil record (Silvestro *et al.* 2015; Clark and Donoghue 2017). Such an extensive lag raises two questions: Firstly, is it plausible to associate two events that are separated by such a long interval of time? And secondly, why did the early diverging lineages of angiosperms (the ANA grade) not undergo a similar radiation?

Schranz *et al.* (2012) proposed a model in which WGD provides latent evolvability that may be later triggered by a shift in environment to promote diversification. This has been further refined and several new models have emerged to explain the lag phase, some of which are readily testable. Among these is the suggestion that it is not WGD, but the ensuing process of genome fractionation (or diploidisation), that may be responsible for diversification. During this process, the organism undergoes large scale genome rearrangements and redundant gene copies are silenced and excised, leading to potentially novel patterns of expression (Dodsworth *et al.* 2016). Most angiosperm lineages have undergone multiple rounds of WGD and exhibit the fastest rate of genome size evolution among land plants (Puttick *et al.* 2015), and it has been proposed that their ability to rapidly downsize their genome in the wake of WGD has led to their global dominance (Simonin and Roddy 2018). Ferns show a higher rate of genome duplication than angiosperms yet

appear not to undergo such extensive genome downsizing and are considerably less diverse than angiosperms (Wood *et al.* 2009; Clark *et al.* 2016). The observed lag between WGD and diversification in angiosperms may be explained by the period of genome fractionation, though the long-term rate of fractionation is uncertain. It seems appropriate to ask whether the extent or rate of genome reorganisation post-WGD correlates with observed shifts in the rate of diversification. The WGD event associated with one of the most dramatic shifts in diversification, the gamma event at the base of eudicots, involved extensive genome reorganization (Jiao *et al.* 2012; Wang *et al.* 2016). Speciation post-WGD would lead to fractionation occurring independently in separate lineages, which could explain the differences between lineages that emerge from WGD (Dodsworth *et al.* 2016).

In the specific case of autopolyploidy (duplication involving a single parental lineage) the newly duplicated paralogs can pair randomly at meiosis. This pattern of tetrasomic inheritance facilitates ongoing exchange between paralogous chromosomes and may prevent them from diverging until a state of disomic inheritance is restored (Martin and Holland 2014; Robertson *et al.* 2017). The period required to attain a state of disomic inheritance could also explain the macroevolutionary lag between WGD and phenotypic evolution. As with the duplication-fractionation model, speciation occurring before the restoration of disomic inheritance will result in independent diploidisation of lineages. (Robertson *et al.* 2017) demonstrated this 'lineage specific ohnolog resolution' (LORe) model in the descendants of the salmonid fish-specific WGD event and showed that independent diploidization was present in 27% of salmonid paralogs. Though untested in plants, its predictions of a long lag period and disparate evolutionary trajectories suggest that it may also fit the patterns observed after the angiosperm-specific WGD.

The case for a general theory of genome duplication as an intrinsic driver of diversification is undermined by the multiple cases where WGD does not accompany

any shift in diversity. Non-seed plant lineages, such as palaeopolyploid mosses and horsetails, remain species-poor despite repeated duplications (Vanneste *et al.* 2015; Devos *et al.* 2016). This can be partly reconciled by the differing rates of genome downsizing and rearrangement exhibited by these clades relative to angiosperms. However, further research on the mechanisms for rapidly altering genome structure are required. Beyond plants and, in particular, among teleost fish, the palaeontological record shows no evidence in support of a role for WGD in directly promoting diversification (Laurent *et al.* 2017). There is some evidence supporting a direct role for WGD in promoting diversity in yeasts where reciprocal gene loss can lead to reproductive isolation (Maclean and Greig 2011), though on a macroevolutionary scale this effect is small (Muir and Hahn 2015).

5 The Origins of WGD

Traditionally, polyploids are recognised as originating from a single parent species (autopolyploidy, xx to xxxx) or from two hybridising species (allopolyploidy, xx + yy to xxyy). Current views maintain that these two outcomes exist along a spectrum, with segmental allopolyploids containing paralogs that display varying levels of synteny (Xiong *et al.* 2011). A segmental allopolyploid may form via hybridisation between two closely related species, or through the process of homoeologous compensation (Xiong *et al.* 2011). Despite potential differences in outcome, both are likely to have had significant effects throughout plant evolution and both processes and their potential evolutionary outcomes have recently been reviewed (Steige and Slotte 2016; Spoelhof *et al.* 2017; Bottani *et al.* 2018). Based on observations from neopolyploids, there is reason to believe that their outcomes may differ, and so it is a priority to establish whether ancient events were a consequence of autopolyploidy or allopolyploidy. Methods to differentiate between the two processes are developing, and in some instances ancient events have been successfully characterised. Genome dominance is a phenomenon observed in allopolyploids, where one subgenome

shows lower expression and retention than the other (biased fractionation). Signal of a bias in gene retention between subgenomes could provide evidence for allo- rather than autopolyploidy (Garsmeur *et al.* 2014). Gene tree methods are also capable of resolving allopolyploid WGDs by considering reticulate patterns of gene tree evolution (Marcet-Houben and Gabaldon 2015; Julca *et al.* 2017) and in some instances they have been able to identify the most likely parental lineages involved in the hybridization event (Gregg *et al.* 2017).

The nature of WGD impacts on the approach required for dating as both auto- and allopolyploidy present different issues. The two subgenomes of an allopolyploid would have diverged at the point of speciation between the two parent lineages, rather than the hybridisation event itself (Doyle and Egan 2010; Kellogg 2016). Successful and viable hybrids are more likely to arise between closely related species, giving rise to 'segmental allopolyploids'. However, there are examples within plants of hybridisation between distantly related lineages (Rothfels *et al.* 2015), which could lead to a significant overestimation of the age of the WGD. Similarly, as outlined previously, autopolyploidy can lead to a prolonged period of tetrasomic inheritance between paralogs (Robertson *et al.* 2017). In this case there is the potential to underestimate the age of the WGD, as the paralogs will only start to diverge once disomic inheritance has occurred, and we date the point at which they diverge rather than duplicate.

6 WGD and Morphological Innovation

The link between WGD and morphological evolution in plants has remained both pervasive and speculative (Ohno 1970; Crow and Wagner 2006). Some have proposed that polyploids may survive and evolve in extreme or marginal habitats, allowing them a competitive advantage over their parent species at range margins (Stebbins 1947). However, the range of many extant polyploids does not exceed that of their parents (Glennon *et al.* 2014), while genes related to stress tolerance appear to have

evolved via tandem duplication rather than WGD (Hanada *et al.* 2008; Panchy *et al.* 2016). The evolution of morphological diversity, like species diversity, may also require a lag phase. For selection to act on innovation, developmental robustness is required (Melzer and Theissen 2016), and so it is possible that morphological diversification may occur only after a period of developmental lability. At the genetic level, WGD may free a lineage from the constraints of purifying selection and allow genes to take on new functions (Ohno 1970). At the phenotypic level this may allow the evolution of novel forms and body plans. Indeed, formative innovations within the plant kingdom have been associated with the expansion of families of regulatory genes (Rensing 2014; Chanderbali *et al.* 2016). Patterns of gene retention post-WGD are not random and in repeated cases genes encoding proteins that function as part of networks and signalling cascades, are retained preferentially (Seoighe and Gehring 2004; Veron *et al.* 2007; Qiao *et al.* 2018). This has been explained in terms of dosage balance and the need to maintain stoichiometric ratios of proteins within the cell (Veitia *et al.* 2008; Birchler and Veitia 2012). The dosage balance hypothesis is exemplified during the diploidisation process in allopolyploids, where exchanges can occur between homoeologous chromosomes of subgenomes (Xiong *et al.* 2011). These exchanges can result in novel gene expression and gene copy number (Lloyd *et al.* 2018), but can also result in the deleterious loss of chromosome regions or entire chromosomes. Homoeologous compensation has been proposed as a mechanism to prevent dosage imbalances and has been demonstrated to lead to increased phenotypic variation in newly synthesized allopolyploids (Xiong *et al.* 2011). The dosage balance hypothesis does not predict the evolution of morphological diversity until such constraints are relaxed and retained paralogues are selected to evolve new functions (Freeling and Thomas 2006; Conant *et al.* 2014). These constraints may relax under different selection pressures although a quantitative model of compensatory drift has also been proposed (Thompson *et al.* 2016). Compensatory drift is the process whereby paralogs are initially retained due to dosage sensitivity, but over

time expression levels of the individual genes drift until one paralog is free of the dosage-dependent constraint (Thompson *et al.* 2016). This model not only provides a mechanism for neofunctionalization to arise from a state of dosage balance, but also a potential explanation for the emergence of evolutionary novelty after prolonged periods of evolutionary time.

It is difficult to ascribe adaptive evolution to WGD, especially with ancient events. The link between WGD and novelty has been elegantly shown in the glucosinolate pathway in Brassicales (Edger *et al.* 2015). This gene family has expanded over several rounds of WGD and is involved in plant-herbivore interactions. It has also been proposed that gene families underpinning floral patterning, expanded during the angiosperm-specific WGD (Chanderbali *et al.* 2016). These genes are implicated in the origin and diversification of the flower, a structure that has shaped recent plant and animal evolution (Fernandez-Mazuecos and Glover 2017). The evolution of pentamerous flowers in the core eudicots also coincides with a genome triplication not shared by basal eudicot lineages (e.g. Proteales, Buxales) (gamma, Fig 2.1) (Jiao *et al.* 2012; Chanderbali *et al.* 2017). The coincidence of the gamma event with this major synapomorphy, a large increase in the rate of diversification, and extensive genome reorganisation (Wang *et al.* 2016), makes it a tantalising system in which to investigate the link between WGD and morphological evolution.

Regulatory gene retention and large shifts in patterns of their transcription suggest a role for WGD in the evolution of eudicot floral diversity (Chanderbali *et al.* 2017). In order to make such a hypothesis testable, the increase in phenotypic complexity must be quantified for comparative analysis (Oyston *et al.* 2016). To achieve this, we can borrow from palaeontology, which has a strong tradition in comparative analysis of phenotype through multivariate statistics – manifest as "morphospace analyses". The hypothesis that WGD drives innovation would predict that events coincide with either the movement to a new 'island' within

morphological design space or a continued expansion of an existing one. These predictions can be tested explicitly with datasets that use discrete morphological characters to describe the traits that unit and distinguish taxa (Hetherington *et al.* 2015). For example, we can characterise the disparity of extant angiosperms to test the hypothesis that the gamma triplication event coincides with an increase in morphological diversity. To do this we used a morphological dataset that captures the disparity of early angiosperms, basal eudicots and core eudicots (Nandi *et al.* 1998). I used these data to calculate the dissimilarity between each taxa, as measured using Gower's dissimilarity metric (Gower 1971). To visualise this dissimilarity, we performed non-metric multidimensional scaling, a non-metric ordination method that summarises variation over a specified number of axes – in this instance, two. The result is presented in Figure 2.3 which shows that the core eudicots occupy a far greater area of morphospace than the basal eudicots. Furthermore, relative to other early diverging lineages of angiosperms, they occupy the largest proportion of morphospace (partial disparity, Fig 2.3b). In addition, we subsampled the character matrix for just floral characters, relating specifically to the gamma-derived hypothesis (Fig 2.3c). The resulting morphospace shows less separation between the lineages, but core eudicots still occupy the largest area and, therefore, exhibit the greatest variation. The construction of a morphospace can be subjective in that it is dependent on the choice of taxa and characters - yet there is strong evidence to suggest that the gamma triplication coincides with the rapid evolution of morphological disparity among eudicots. A comparable analysis of the impact of WGD in Pines finds support for increased variance in morphospace occupation, but gross uncertainty in the estimate of the timing of WGD relative to the age of the disparate clade undermines the hypothesis of a causal link.

Quantifying morphological evolution across multiple lineages will be instrumental to understanding the role of WGD in the evolution of phenotypic complexity. The inclusion of fossil taxa and recent methods used to estimate

disparity through time may allow us to measure the tempo of morphological evolution post-WGD. The impact of key innovations that are attributed to WGD can be tested by considering their impact on the shape of a morphospace or whether the innovation has resulted in diversification. A further question arises as to what degree WGD is essential for the appearance of major innovations. For example, the origin of seed and flowering plants coincides with a WGD event yet, arguably, a greater number of characters unite the vascular plants whose origin was independent of any known WGD events (Banks *et al.* 2011). While it is plausible that saltational evolution has been effected by WGD in the plant kingdom (Minelli 2018), phenotypic complexity may also arise through the evolution more nuanced trans- and cis-acting regulation (Chen *et al.* 2012).

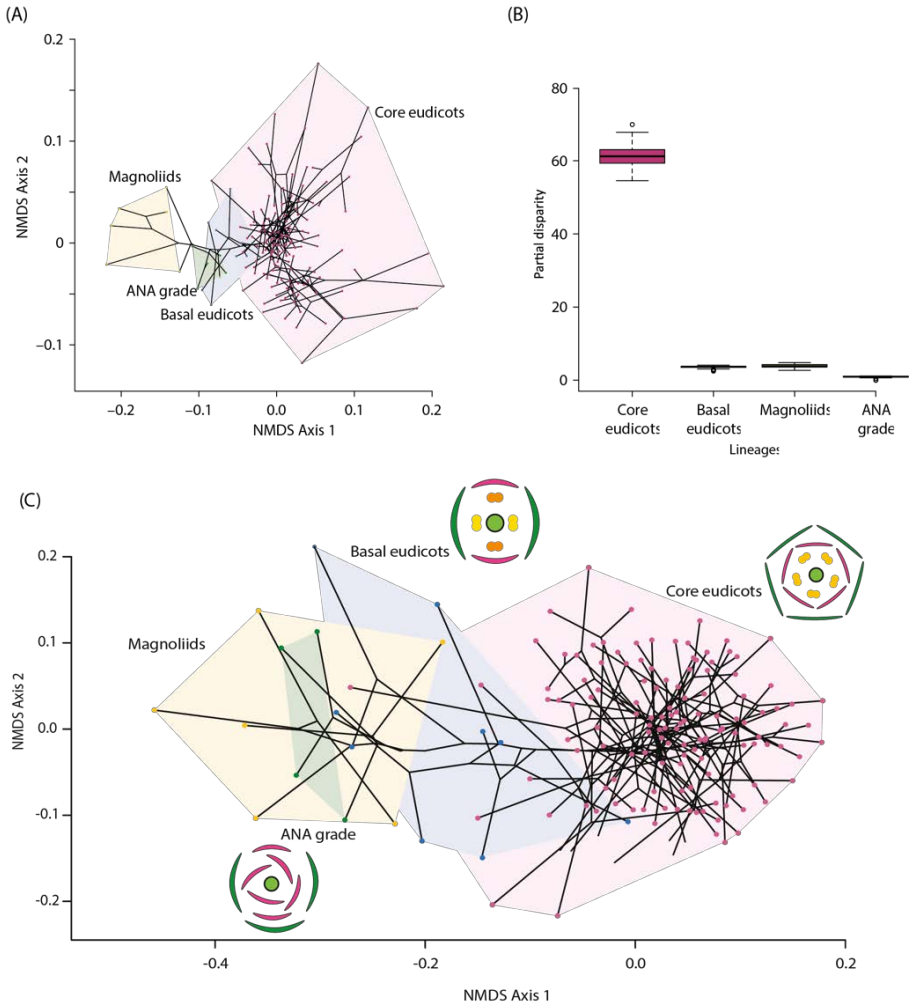


Figure 2.3. Morphological evolution in the wake of the gamma triplication which occurred before the evolution of the core eudicots. A) an empirical morphospace based on a morphological matrix (Nandi *et al.* 1998). Morphological characters form the basis of a distance matric (Gower's Index) which is subjected to non-metric multidimensional scaling (NMDS) to display variation in two axes. A consensus phylogeny is mapped onto the morphospace. B) the contribution to total disparity (partial disparity) of each clade calculated from distance matrix (1000 bootstrap replicates) (Guillerme 2015). C) A morphospace constructed from the floral characters. Major trends in floral evolution are displayed next to the lineages, with spiral phyllotaxis present in early angiosperms, the dimerous flowers common among basal eudicots and the pentamerous flower that is associated with the core eudicots.

6.1 Duplication and Disparity in the Conifers

Some explosive genome duplication events, such as those associated with the core eudicots, coincide with rapid diversification and an increase in morphological variation. However, many WGD events in species-poor lineages are not closely associated with a macroevolutionary phenomenon. Most conifers are thought to have undergone at least two rounds of WGD during their evolution, one shared among seed plants and then two lineage-specific events on the branches leading to Pinaceae and Cupressophytes (Li *et al.* 2015). Preliminary analyses of diversity and disparity in the pines indicate a rapid increase in morphological variance during the late Jurassic and Early Cretaceous (Oyston *et al.* 2016) and Pinaceae occupies a highly distinct area of morphospace (Fig 2.4). This provides some corroborative support for the hypothesis that WGD has resulted in morphological variation among conifers during their early evolution. However, the age of the pine WGD is currently estimated between 342 and 200 Ma (Li *et al.* 2015) (Fig 2.4); with so much uncertainty it is not presently possible to link WGD to the shift in morphological disparity. This example highlights the need to employ methods that can accurately and precisely estimate the timing of WGD events as a temporal framework is essential for testing macroevolutionary hypotheses (Clark and Donoghue 2017).

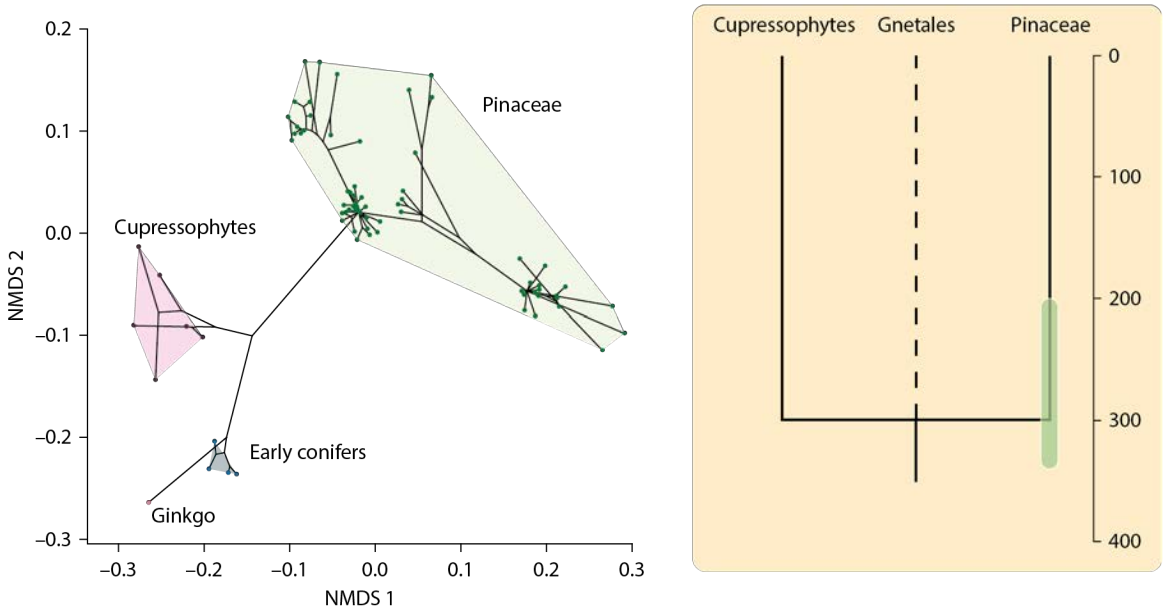


Figure 2.4. Morphological evolution in the Pinaceae. An empirical morphospace of Pinaceae and relatives built from morphological characters (Smith *et al.* 2017) which formed the basis of a distance matrix (Gower's Index) that was subjected to NMDS. A consensus phylogeny is mapped onto the morphospace. The uncertainty of both the relative (phylogenetic) and absolute timing (green bar) of the event limits our understanding of the consequences since the position of the Gnetales remains contentious and the current estimate for the age of the WGD spans over 100 Myrs.

7 Conclusions

WGD is associated with a macroevolutionary outcome in some, but not all lineages, and it remains unclear how and why is this the case. As the number of identified WGD events in plant evolutionary history increases, there is an ever-growing need for a general theory on the role of WGD in macroevolution. However, in order to establish whether WGD is a class of event with characteristic and predictable outcomes, further work is needed in order to place, both relatively and absolutely, each event in time. There are many outstanding questions to be answered, but a precise temporal framework forms the basis for tests that can quantify any macroevolutionary consequences and inform and refine hypotheses about the

relationship between WGD, diversification, and morphological evolution. Plants are the best system in which to elucidate the effects of WGD because of the prevalence of these genomic events in plant phylogeny. This will be crucial as we seek to explain the consequences beyond any single event and, given the role that genome duplication has had in the evolution of many crop species, being able to make general predictions about the outcome of WGD is of critical interest.

Chapter 3

Constraining the timing of whole genome duplication in plant evolutionary history

1 Summary

Whole genome duplication (WGD) has occurred in many lineages within the tree of life and is invariably invoked as causal to evolutionary innovation, increased diversity, and extinction resistance. Testing such hypotheses is problematic, not least since the timing of WGD events has proven hard to constrain. Here we show that WGD events can be dated through molecular clock analysis of concatenated gene families, calibrated using fossil evidence for the ages of species divergences that bracket WGD events. We apply this approach to dating the two major genome duplication events shared by all seed plants (ζ) and flowering plants (ε), estimating the seed plant WGD event at 399-381 Ma, and the angiosperm WGD event at 319-297 Ma. These events thus took place early in the stem of both lineages, precluding hypotheses of WGD conferring extinction resistance, driving dramatic increases in innovation and diversity, but corroborating and qualifying the more permissive hypothesis of a 'lag-time' in realising the effects of WGD in plant evolution.

This chapter has been published as Clark, J.W. & Donoghue, P.C.J. 2017 Constraining the timing of whole genome duplication in plant evolutionary history. *Proceedings of the Royal Society B: Biological Sciences*, 284 (1858) . *The idea was conceived by P. Donoghue and the author, the experiments were designed by and performed by the author and the article was written by the author.*

2 Introduction

The discovery in plant genomes of evidence of recurrent whole genome duplication events (WGD; polyploidy) has reignited debate over its importance in land plant evolution (Soltis *et al.* 2014b; Mayrose *et al.* 2015). Several causal hypotheses have emerged linking WGD to key innovations (Soltis and Soltis 2016), increased rates of diversification (Tank *et al.* 2015) and extinction resistance that may have facilitated the success of multiple lineages of extant plants (Vanneste *et al.* 2014b). The mechanisms through which genome duplication can result in evolutionary novelty are becoming better understood and the traditional models of neo- and subfunctionalisation have now been hybridised with models of dosage balance in attempts to explain how evolutionary innovation can arise post-WGD in the face of extensive gene loss and stabilising patterns of gene retention (Conant *et al.* 2014; Teufel *et al.* 2016). Furthermore, there now exist elegant examples of genes and gene families that have taken on new functions (neofunctionalisation) following multiple rounds of WGD and then playing a key role the evolution of plant lineages (Edger *et al.* 2015). The link between polyploidy and diversification remains controversial (Kellogg 2016), but there exists some evidence that several of the ancient WGD events in angiosperms correlate with shifts in diversification (Tank *et al.* 2015). Separating the WGD events and the shifts in diversification are a 'lag' of several million years, which has been explained as the period of fractionation post-WGD and, in turn, the feature of WGD that leads to innovation and diversification (Dodsworth *et al.* 2016). However, at the broadest scale, these hypotheses are underpinned by the relative phylogenetic placement and absolute timing of each event. Though the relative phylogenetic timing of plant WGD events is well constrained, their absolute timing is not (Kellogg 2016).

Constraining the phylogenetic position of WGD events relies on broad taxonomic sampling of genomic or transcriptomic data. The presence or absence of shared 'age peaks' in *Ks* plots of synonymous substitution rates between duplicates

provides evidence for shared genome duplications (Lynch and Conery 2000). This approach culminated in a survey of 41 plant genomes focussing on angiosperms (Vanneste *et al.* 2014b) and more recently several transcriptomes also highlighting the presence of WGD within the evolutionary history of gymnosperms (Li *et al.* 2015) and peat mosses (Devos *et al.* 2016). The number and position of the peaks on the *Ks* plot also reveals the relative timing of each event, with multiple peaks representing multiple successive WGDs. The absolute timing of each event can be obtained indirectly by phylogenetically bracketing the event – the event must have occurred along the branch between those lineages that have undergone the WGD and those that have not. However, despite well-sampled exceptions among certain groups of angiosperms (Estep *et al.* 2014; Kagale *et al.* 2014; Barker *et al.* 2016), there are few cases where the sampling of taxa is dense enough to prevent very long branches, and so the ages of genome duplication events must be inferred directly. Direct dates can be obtained by converting the relative timing of peaks on a *Ks* plot into absolute ages. This has the advantage that it does not require additional taxon sampling and so estimates can be obtained for WGD events isolated on long branches (Vanneste *et al.* 2015). A major caveat of this approach is that it relies on the assumption of a strict molecular clock that, depending on shifts in the rate of sequence evolution, can lead to inaccurate age estimates. Furthermore, *Ks* plots are known to saturate beyond a certain age, meaning that they cannot always distinguish more ancient duplications and may lead to artificial peaks in the distribution (Vanneste *et al.* 2013). More complex relaxed clock methods can be employed in a phylogenetic or phylogenomic approach, whereby the individual gene families containing signal of WGD are reconstructed and individually dated (Fawcett *et al.* 2009). The distribution of ages obtained can then be plotted to provide a range of estimates for each event. This approach is more powerful and has been used to estimate the ages of multiple WGD events across the angiosperms, where genomic and transcriptomic data are more abundant (Fawcett *et al.* 2009; Jiao *et al.* 2011). However, dating individual gene trees

does not fully exploit the power of the molecular clock, and the power of individual gene trees is likely to diminish over longer periods of evolutionary time. Increasing the amount of sequence data by concatenating multiple gene families into alignments decreases uncertainty in the estimation of relative ages (dos Reis *et al.* 2016), and can be used to date the absolute timing of WGD events (Macqueen and Johnston 2014) yet, to date, studies focussing on WGD in plants have relied on the power of individual gene trees. Directly dating WGD events using concatenated gene trees also provides estimates of the absolute timing of the WGD in relation to subsequent speciation events within the lineage, since gene trees observe species divergences as well as duplication events. Thus, concatenated gene trees have the potential to provide an accurate estimate of the absolute timing WGD events relative to the diversification events in which they are causally implicated.

The seed plants (Spermatophyta) are the most species rich of extant plant clades, encompassing the gymnosperms and angiosperms (flowering plants). WGD events have been identified at the base of all seed plants (ζ ; Jiao *et al.* 2011; Li *et al.* 2015) and at the base of all angiosperms (ε ; Jiao *et al.* 2011), and so all extant flowering plants have undergone at least two rounds of genome duplication. Previous attempts to date these events were based on distributions of ages inferred using poorly defined calibrations and penalized likelihood molecular clock methods (Jiao *et al.* 2011) that have since been found unreliable (Thorne and Kishino 2005). The WGD shared by all extant angiosperms has been linked with the 'big bang' diversification of the Mesangiospermae (following a lag period) as well as several major innovations, including the origin of the flower (Tank *et al.* 2015; Soltis and Soltis 2016). WGD has been thought to be less prevalent within gymnosperms, the sister clade to angiosperms (together comprising Spermatophyta), despite the fact that the ζ WGD is part of their shared evolutionary history. More recent evidence has indicated that WGD has occurred in several gymnosperm lineages and confirmed that the ζ WGD (Spermatophyte) was not shared with their sister lineage, the ferns (Li *et al.* 2015).

Conventionally molecular clock dating approaches have sought to minimise the influence of duplication by using only single copy genes. In contrast, we exploit the pattern of paralogy produced by WGD in the evolutionary history of multiple gene families and concatenate them into a partitioned alignment. Combined with broad taxon sampling and multiple fossil calibrations, we demonstrate an approach for dating gene trees to provide well-constrained estimates of the timing of duplication events and attendant speciation events.

3 Materials and Methods

Gene families containing signal of the ζ (spermatophyte) and ε (angiosperm) WGD events and those that contain the signal of both were catalogued by Jiao et al. (2011), and from these we expanded orthogroups by obtaining amino acid sequences using Plaza 3.0 (bioinformatics.psb.ugent.be/plaza), and GreenPhyl 4 (www.greenphyl.org). Further sequences were obtained by local BLAST searches of iPlant (*www.iplantcollaborative.org*). 128 species were sampled in total, representing all major lineages of land plants and these are listed in Appendix 1. Four datasets were assembled for all taxa: families containing a clear signal of just the ε WGD event (angiosperm dataset), just the ζ WGD event (spermatophyte dataset), families containing signal of both events (ζ+ε dataset), and a combined dataset. To verify a clear signal of the relevant WGD event in each gene family, we built individual gene trees based on multiple amino acid sequence alignments generated using MAFFT while model selection and gene tree reconstructions were performed using iQtree (Nguyen *et al.* 2015). We opted for a conservative approach, discarding orthogroups that following phylogenetic reconstruction and visual inspection did not clearly reflect the signal of either or both WGDs (Fig 3.1), had sequence alignments shorter than 100 amino acids, displayed a topology that was incongruent with our current understanding of land plant phylogeny with either the total group seed plants or major lineages within being resolved as non-monophyletic, or were too large with

multiple nested duplications, resulting in large numbers of sequences having to be discarded.

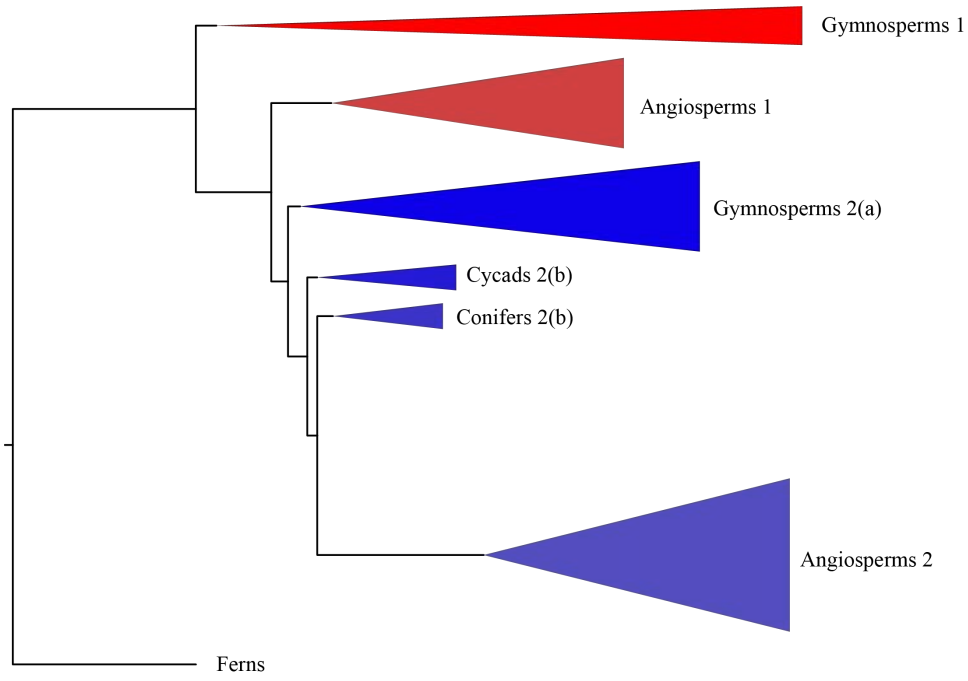


Figure 3.1. An example of an orthogroup that was discarded from the analysis. Triple-Helix transcriptor family (ORTHO03D004565) was identified by Jiao *et al.* (2011) as containing the signal of the *zeta* duplication. Though not rejected, the signal is difficult to recover, due to an incongruent topology (paraphyletic gymnosperms) and the relationships of the two sets of paralogs not being clear.

Of 130 orthogroups surveyed, 12 gene families were found containing a clear signal of the ε WGD. The number of sequences among individual gene families ranged from 87-126 and when concatenated a total of 176 tips. 14 further gene families were found for the ζ WGD, representing 189 tips when concatenated and varying from 106-149 tips individually. An additional 7 gene families were found containing the signal for both, for which 254 tip sequences were assembled when concatenated and individual gene families ranging from 132-249 tips. The combined dataset contained 33 gene families, with one node representing ζ, but two representing ε. As 12 gene families

contain only one node with the ε duplication, the event was represented only once in the combined analysis, to maximise precision at this node. Similarly, angiosperm gene copies from gene families not containing signal of the ε duplication were randomly assigned to one side of the duplication. Due to differential retention, a copy of each gene paralog was not present in all families and the number of tips in each gene varies.

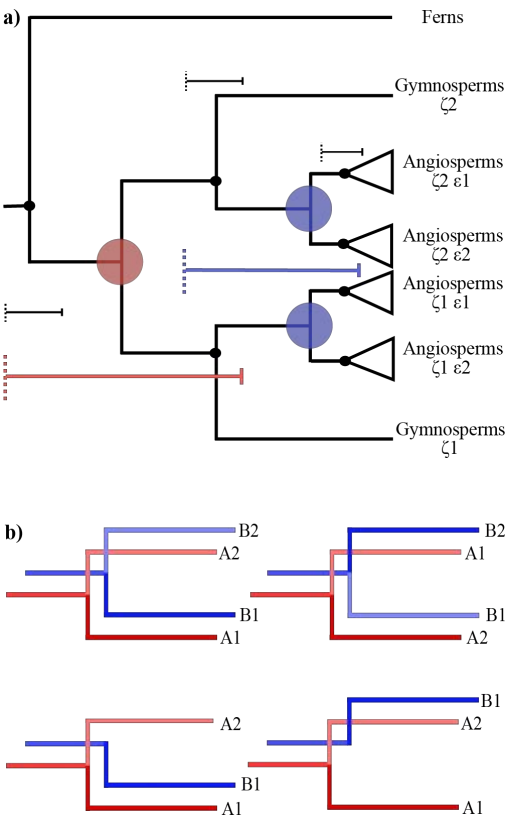


Figure 3.2. a) An example gene tree showing the seed plant (ζ, red) and angiosperm (ε, blue) duplications. The duplication events are constrained using minima and maxima (coloured brackets) based on fossils used to constrain speciation events (black brackets). b) Gene trees may retain both copies of the duplicate gene (top row), or a single copy may be lost (bottom row). When concatenating duplicates from different gene families, given that both copies are descended from the same event, their assignment to either side of the duplication is arbitrary.

Across all analyses, nodes were constrained using 35 fossil calibrations spanning land plant phylogeny defined using best practice (Parham *et al.* 2012) Appendix 2). The duplication nodes were constrained temporally to reflect the possibility of the WGD occurring at any point following the divergence of spermatophytes from an ancestral euphyllophyte (ζ WGD event) and for angiosperms from an ancestral spermatophyte (ε WGD event; Fig 3.2). Calibrations that provided only a minimum age were modelled as a hard minimum bound with a truncated Cauchy distribution (p = 0.1, c = 0.2). Calibrations that provided a maximum age were

modelled with a soft maximum with a uniform distribution between the minimum and maximum age (Warnock *et al.* 2015). Molecular clock analyses were conducted on concatenated alignments using the normal approximation method in MCMCTREE under the appropriate model (Yang 2007). The normal approximation method provides a fast and efficient way of analysing large datasets using complex models and a relaxed clock and is run under a fixed topology. We ran all analyses on a topology reflecting both WGD events and recent hypotheses of relationships among land plants (Wickett *et al.* 2014) (Appendix 2). We also reconstructed the topology based on our own datasets using iQtree and found that it was highly congruent with the constraint tree. Each analysis was run twice independently and regularly checked for convergence and for effective sample sizes greater than 200 using Tracer v.1.6 (Rambaut *et al.* 2014).

Assuming autopolyploidy, each WGD event produces two daughter nodes that are created simultaneously and that must have the same age, and so the assignment of each paralog to either node of the duplication is arbitrary (Fig 3.2). In this way paralogs between the gene families can be concatenated in multiple combinations, so long as they are consistent within each gene family. To explore the impact of different combinations of paralogy groups between gene families, We randomly reassigned groups to either node using the ζ+ε dataset containing both duplications.

The extent to which the low number of available gene families impacted on the estimation of dates was explored through infinite sites analyses (Yang and Rannala 2006). The gene families were successively concatenated, and the analysis repeated with one more gene family each time. The relationship between the mean age estimates and the widths of the 95% Highest Posterior Densities (HPDs) was used as a measure of the precision of the data versus the uncertainties induced by the fossil calibrations. Higher r^2 values indicate that large HPD widths are due to increasing uncertainty in the fossil record deeper in time. A saturation of the curve

suggests that adding further sequence data would not increase the precision of the analysis, since it is limited by the information available in the fossil record.

4 Results

In most Bayesian molecular software, specified node age priors are modified in the construction of the joint time prior to achieve the expectation that only ages compatible with the assumption that ancestral nodes are older than their descendants, are proposed to the MCMC (Inoue *et al.* 2010; Warnock *et al.* 2012). To ensure that these effective priors are biological reasonable, we estimated them by running the analysis without sequence data. The effective priors are compatible with the original palaeontological and phylogenetic evidence, yielding broad 95% HPDs for the timing of WGDs in all analyses, though both were truncated relative to the specified calibrations. The spans of the 95% HPD for the prior on the ζ and ε WGD events are 81 (434-353 Ma) and 111 (355-244 Ma) million years, respectively (Table 3.1). In the separate analyses of both the ζ and the ε WGD events, the truncation effects on the prior were the same as for the combined analysis, and so the additional nodes in the combined analysis and the ζ+ε dataset did not affect the effective prior.

In all instances, the addition of sequence data yielded estimates congruent with, yet more precise than, the joint time prior. Estimates for both WGD events were compared between gene families using the ζ+ε dataset, and we found variation in both the width of the 95% HPD and the absolute age estimates, though the overlapping distributions of the HPDs showed that the gene families were congruent. While some gene families produced much more precise estimates, the variation in estimates between all gene families showed a similar level of precision to the joint time prior alone, ranging from 435-346 Ma for the ζ WGD event and 355-244 for the ε WGD event. The ζ+ε dataset also allowed us to compare the estimates for the ε duplication, which is represented twice in each gene family, within gene families. We found that the 95% HPD widths for the event varied within gene

families, though this is likely due to the absence of paralogs on one side of the duplication. The only family with all paralogs present, CDK, showed estimates consistent in both age and uncertainty across both nodes. The greatest effect in terms of precision was produced by increasing the amount of sequence data by concatenating the gene families. The effect of missing paralogs across both duplication nodes in the ζ+ε dataset was minimised and the age estimates for both ε nodes were highly consistent. The ζ+ε concatenation was also considerably more precise than any of the individual gene families (Table 3.1). Multiple concatenations were tested on this dataset, to determine if the assignment of paralogs between duplicates affected the estimates. We did not observe any material differences in age or uncertainty, indicating that the results are robust to the way in which the gene families are concatenated.

The addition of further sequence data for each duplication event in turn produced results of even greater precision. The angiosperm dataset estimated an age of 321-295 Ma for the ε WGD event, almost 5 times more precise than the joint time prior alone. A similar increase in precision was obtained by the spermatophyte dataset, the ζ duplication estimated to have occurred 400-380 Ma, 4 times more precise than the joint time prior alone. Based on the largest amount of data, the combined analysis of the combined dataset produced results that were highly congruent with the two individual datasets, estimating 399-381 Ma and 319-297 Ma for the ζ and ε WGD events, respectively (Fig 3.3).

Table 3.1. 95% Highest Posterior Density (HPD) estimates for the age of both WGD events, summarising the effective prior, individual gene families (1 to 7), the effects of concatenating gene families, the expanded and combined datasets

		Gene families							Concatenated gene families						ε dataset	ζ dataset	Combined dataset
Node	Effective prior	1	2	3	4	5	6	7	1-2	1-3	1-4	1-5	1-6	1-7			
Spermatophyte duplication (ζ)	353 - 434	382 - 435	346 - 411	346 - 411	354 - 418	354 - 404	357 - 415	355- 433	390- 433	386 - 430	380 - 418	380 - 416	377 - 408	378- 409	-	380- 401	381-399
Angiosperm duplication (ε) ζ1 ε	244 - 355	270 - 339	250 - 353	248 - 328	280 - 354	258 - 340	249 - 351	254- 356	273- 336	268 - 323	280 - 323	285 - 331	282 - 325	281- 323	295- 321	-	297-319
Angiosperm duplication (ε) ζ2 ε	244-355	267- 340	273 - 344	247 - 350	245 - 349	277 - 362	247 - 313	245- 355	278 - 333	276 - 330	276 - 322	289 - 338	283 - 325	276- 321	-	-	-

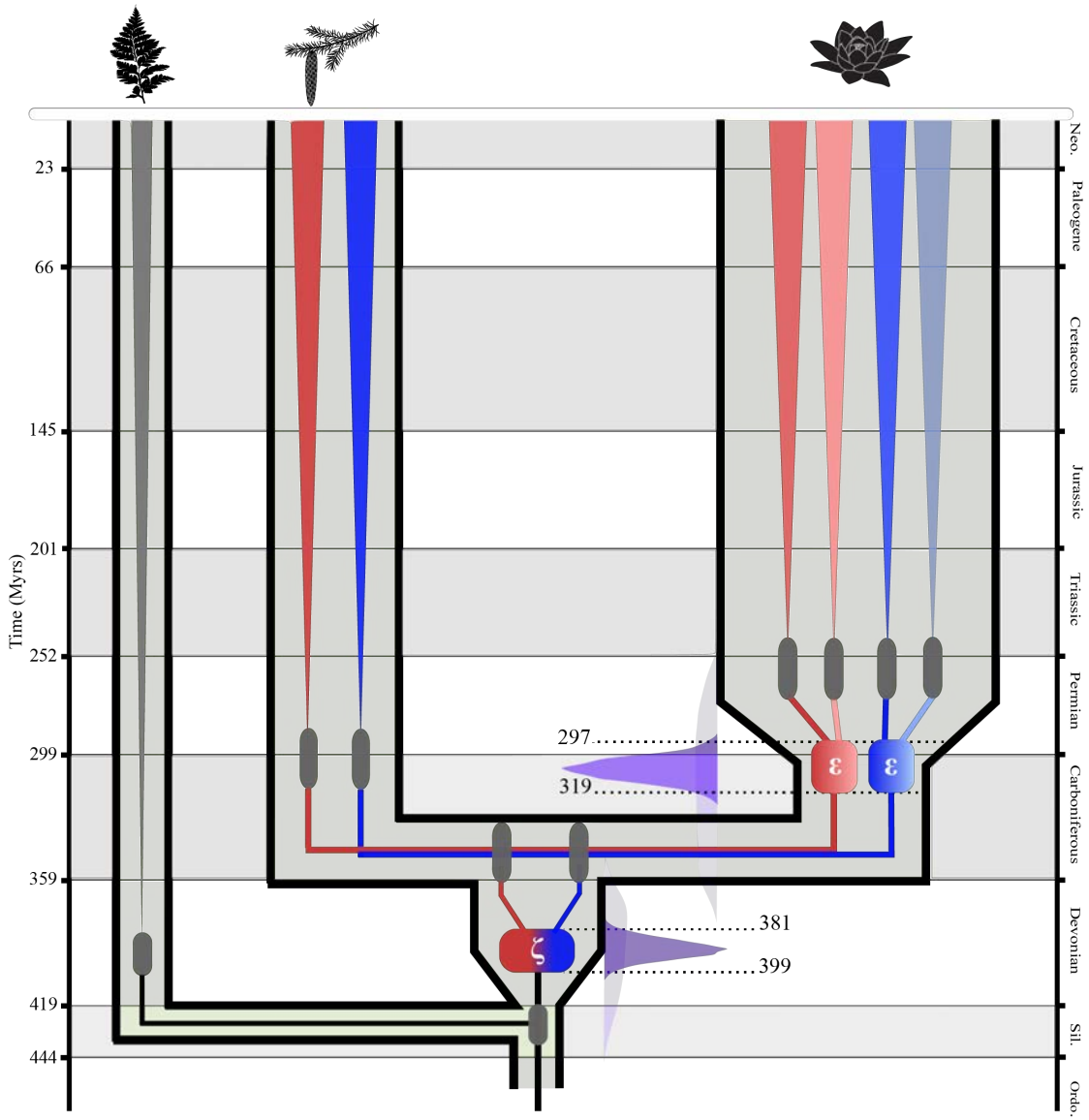


Figure 3.3. Estimated dates for the occurrence of both the seed plant (ζ) and angiosperm (ε) duplication events based on a molecular clock analysis of 33 concatenated gene families. Age estimates (95% Highest Posterior Density) for the divergences of the major lineages and crown groups represented by grey bars. The age estimates (95% HPD) of two duplication events are represented by boxes, with the subsequent subgenomes represented first by blue and red (ζ), then by lighter and darker shades of each colour (ε). For each duplication event, the effective prior is shown (light blue) next to the posterior distribution (dark blue).

Infinite sites plots suggest that though the R^2 value showed little changed with increased sequence data, the addition of sequence data reduced the uncertainty of estimates (Fig 3.4). With 19 gene families, the amount of error was continuing to decrease, suggesting that additional gene families may increase precision further.

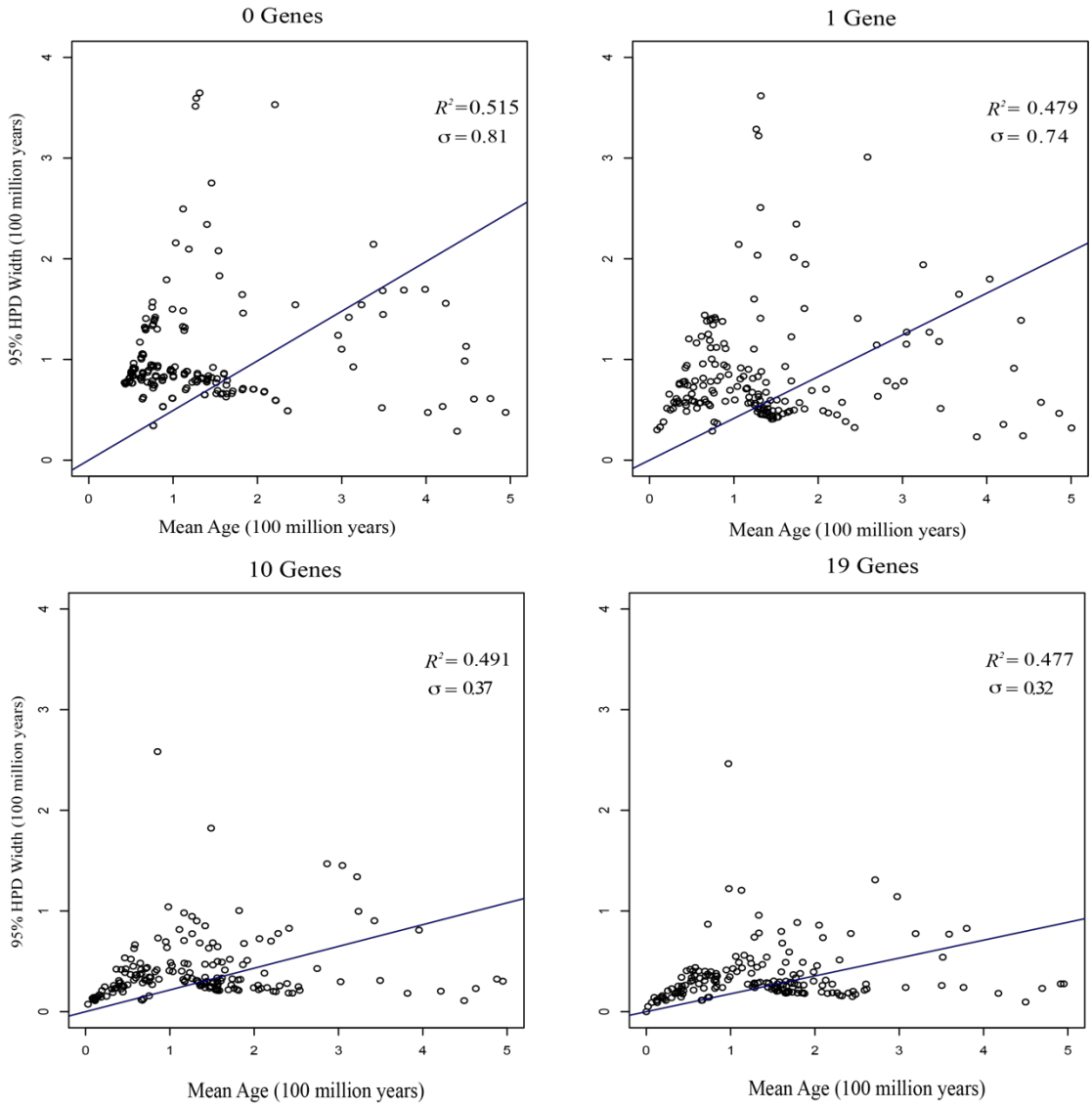


Figure 3.4. Infinite sites plots for the most complete (Angiosperm) dataset, with the regression between the mean age and the 95% Highest Posterior Density shown for 0, 1, 10 and 19 gene datasets. The R-squared and error terms are also shown.

5 Discussion

5.1 Inferring the age of whole genome duplication

Our results indicate that the evolutionary history of gene families can be exploited to obtain precise estimates of the age of WGD events. These methods depend on both careful selection of fossil constraints and available gene families containing signal of WGD events, though even with limited sequence data, We greatly improve the precision over the raw calibrations alone.

Both the ε (angiosperm) and ζ (spermatophyte) genome duplication events have been independently reported ((Jiao *et al.* 2011; Li *et al.* 2015), yet we were unable to find large numbers of gene families with clear signal of either or both events. The paucity of available gene families for these WGD events is likely in part a result of our conservative criteria in selecting gene families based on topology. In part, this reflects the limitations of single genes to resolve unequivocal phylogenetic signal for such events over long timescales. However, it also reflects the antiquity of the events, given that retention of genes following a WGD follows a decay pattern and widespread gene loss leads to a gradually decreasing phylogenetic signal over time. It is unsurprising that so few gene families remain with a clear signal of these events and, when considered next to existing evidence for these events ((Jiao *et al.* 2011; Li *et al.* 2015), our findings are entirely compatible with the ε and ζ duplication events. Our results indicate that the evolutionary history of gene families can be exploited to obtain precise estimates of the age of WGD events. Infinite sites plots lead us to expect that the addition of further sequence data will leverage further precision. Similarly, WGD events that are more recent and may contain more genome-wide data, may be dated using the same approach but with greater precision.

Unlike genomic datasets that can be used for gene-tree reconciliation and the construction of *Ks* plots, the methods presented here focus solely on the dating of WGD events, rather than their characterisation. However, the congruence of age estimates between gene families serves as a test of their coincidence, as anticipated by WGD. The

annotation of gene families to either side of the duplication event requires greater care and is a potentially limiting factor on the number of gene families that can be analysed, yet we have demonstrated that even with a relatively small dataset (compared to a genomic dataset), high levels of precision can be achieved. Novel molecular clock approaches such as cross bracing could also be used to increase precision around the duplication nodes, especially as they are so difficult to constrain (Shih and Matzke 2013).

An additional caveat is that WGD or polyploidy is often categorised into two distinct classes (Garsmeur *et al.* 2014), autopolyploidy and allopolyploidy, traditionally distinguished based on the number of parent species, but also characterised by the patterns of fractionation post-WGD. The mode of duplication may impact our estimates of duplication age (Doyle and Egan 2010), as the point at which duplicates coalesce is actually the timing of divergence of the two parental species, or a more ancestral autoploidy event, as opposed to the alloploidy event itself (Doyle and Egan 2010). New methods are emerging to discriminate between auto- and allopolyploidy (Marcet-Houben and Gabaldon 2015), but these are likely to fail when applied to more ancient genome duplication events. However, allopolyploidy would only have a large impact on accuracy if hybridisation occurred between very distant parent species.

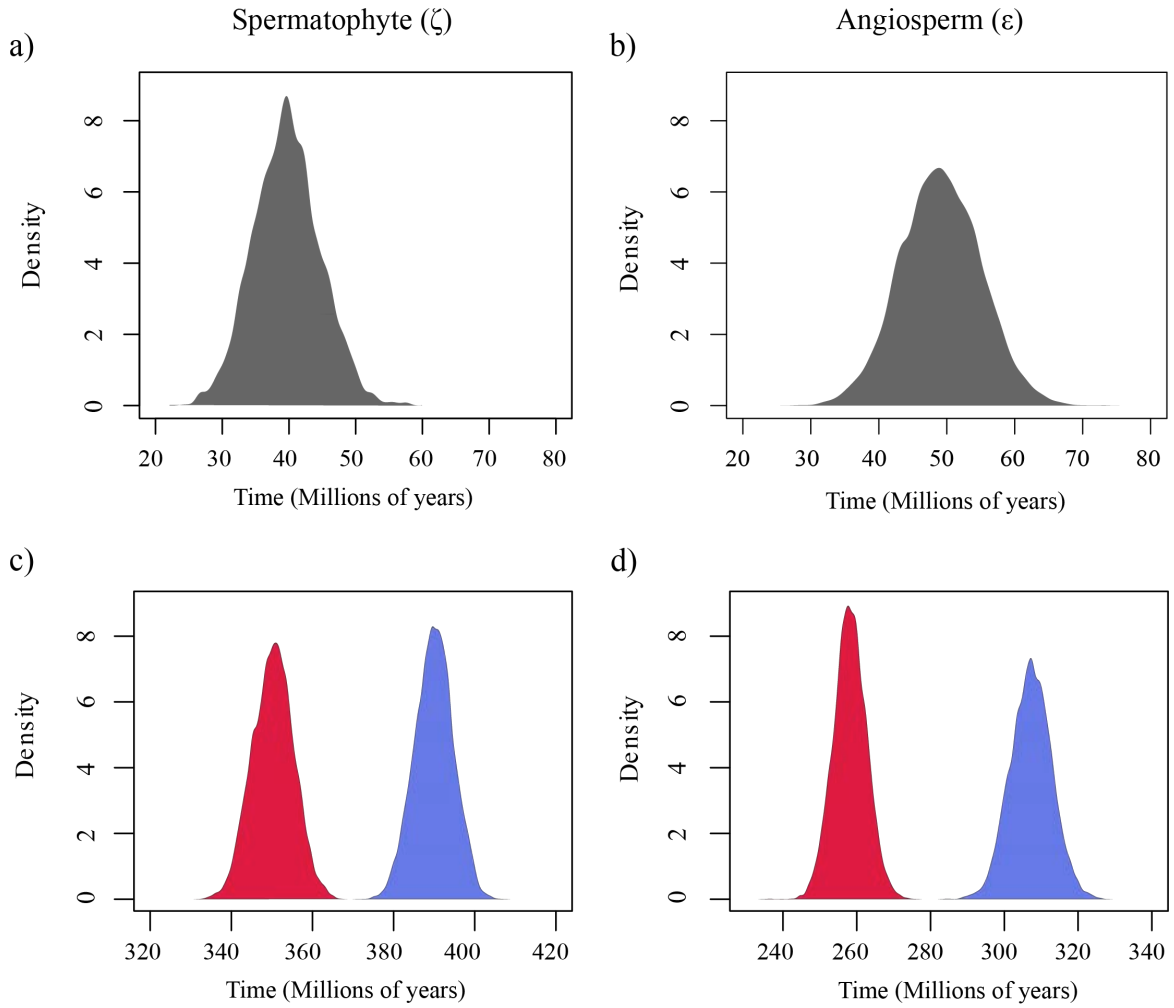


Figure 3.5. The posterior probabilities of a) the lag between the ζ duplication and the diversification of crown Spermatophytes and b) the lag between the ε duplication and the diversification of crown angiosperms. The posterior probabilities of the absolute age of the WGD events (blue) and diversification (red) is also shown for c) ζ and Spermatophytes and d) ε and angiosperms.

5.2 Dating duplication, diversification and innovation

Our most comprehensive analysis of 33 gene families indicated that the genome duplication present in all crown Spermatophytes occurred 399-381 Ma, a period spanning the Early to Late Devonian (Fig 4.3). The WGD event present in all crown angiosperms occurred almost 100 million years later, 319-297 Ma, across the Carboniferous-Permian boundary (Fig 3.3). Gene trees contain both the signal of WGD and species divergence, allow a direct estimation

of the age of the WGD event relative to the age of the crown group (Fig 3.5). Both estimates predict that the respective WGD events occurred early in the stem of both lineages, predating the diversification of the crown group by about 50 million years. These estimates are considerably older than those of Jiao *et al.* (2011), yet our estimates for the age of the seed plant (360 - 340 Ma) and angiosperm (267 – 247) crown groups are comparable to other molecular clock analyses (Foster *et al.* 2017; Murat *et al.* 2017), allowing us to reject the notion that the duplications occurred late in the stem lineage. Greater precision in the absolute age of WGD events leveraged by concatenation allows that hypotheses can be more rigorously tested. WGD occurring early in the stem lineage has two implications for current hypotheses regarding the role of WGD in plant evolution.

First is the hypothesis that WGD drives evolutionary success (Arrigo and Barker 2012; Madlung 2013; Soltis *et al.* 2014a), or confers extinction resistance (Fawcett *et al.* 2009; Fawcett and Van de Peer 2010), since the long stem lineages of both groups are the result of the extinction of many lineages. However, many of these extinct lineages must also share these genome duplications. For example, the ζ duplication predates the appearance of the earliest seed plants, the pteridosperms and Cordaitales, and so WGD cannot have conferred extinction resistance, as has been proposed for the ancient palaeopolyploid *Equisetum* (Vanneste *et al.* 2015). The long-term evolutionary success of seed plants and especially angiosperms is unquestionable, and there is considerable evidence for the role of gene duplication in the evolution of angiosperms, in particular (Chanderbali *et al.* 2016; Soltis and Soltis 2016), yet our results are more in keeping with the idea of 'rarely successful polyploids' (Arrigo and Barker 2012). The challenges faced by polyploids in order to establish and persist may be partially responsible for extinctions in a lineage post-WGD, and it may be the case that extant Spermatophytes and angiosperms are the surviving lineages best able to exploit any long term competitive advantages (Fawcett and Van de Peer 2010). Secondly, if the crown clades of seed and flowering plants can be considered to be characterised by evolutionary success, this has been achieved in both lineages after a substantial lag post-WGD. Our results indicate that the lag between the ζ WGD event and the divergence of

crown Spermatophytes is 22 - 60 million years, and 27 - 65 million years between the ε WGD event and the divergence of crown angiosperms (Fig 3.4). These are comparable to the results of Tank *et al.* (2015) who estimated a 49.2 million year lag between the ε WGD event and the shift in diversification of angiosperms, though without directly inferring the age of the WGD. Tank *et al.* (2015) also estimated that the rate shift in diversification among angiosperms occurred at 213 Ma, following the divergence of Mesangiospermae which, following our age estimates, indicates a lag of 84-106 million years. Ultimately, these results indicate that more precise age estimates require more precise hypotheses regarding the role of WGD in promoting evolutionary success. Given these long lag periods and that some, though clearly not all, clades that share a history of WGD are diverse or characterised by innovations, it requires more explicit hypotheses regarding which clades are considered successful.

Evidently, we find no direct support for the deterministic role of WGD in driving diversification or innovation. Rather, our data are more compatible with the more permissive model of evolution via genome duplication that emphasises the importance of the post-WGD period of genome fractionation. During this period, the need to maintain a dosage balance of protein products selects for the maintenance of duplicates, followed by a relaxation of selection allowing sub- and neofunctionalization (Conant *et al.* 2014). An additional consideration is the Lineage Specific Re-diploidization model, which applies when species divergence occurs before the diploidization process in complete (Robertson *et al.* 2017). Under this model, the lag is produced by the pattern of tetrasomic inheritance that is characteristic of autopolyploidy, leading to massively delayed functional divergence of duplicate genes. This model also predicts that duplicate genes evolve independently in separate lineages, and that this can explain the divergent evolutionary trajectories of lineages that share the same history of WGD (Robertson *et al.* 2017). This more permissive model explains the 'long fuse' or 'lag' found in our results, whereby an early WGD during a lineages evolution provides a primer for subsequent innovation and diversification, leading to the evolutionary success of both lineages (Fawcett and Van de Peer 2010). It also explains

the paucity of genes preserving all paralogues anticipated as a phylogenetic footprint of the ζ and ε WGD events, as a consequence of post-duplication dysploidy leading to dosage bias.

The quantification of this lag is clearly relevant to understanding the role of WGD in plant evolution (Fawcett and Van de Peer 2010). Our methods are applicable to other WGD events characterised previously within the plant kingdom, including those thought to be associated with increased diversification or the K-Pg boundary (Vanneste *et al.* 2014b; Tank *et al.* 2015). Furthermore, these methods could be used to clarify the timing of the proposed WGDs associated with the origins and early evolution of vertebrates (Donoghue and Purnell 2005), which are still undermined by uncertainty around their timing.

6 Conclusions

Accurate and precise estimates of the timing of WGD events are fundamental to our understanding their significance on a macroevolutionary scale and can be achieved by coupling a careful appraisal of the fossil record with molecular clock approaches. We demonstrated that by concatenating multiple gene families with a shared history of WGD into a single alignment, the ages of two ancient WGD events, ε (angiosperm) and ζ (spermatophyte), were estimated to a high degree of precision. Both events were found to occur early in the stem of each lineage, predating the divergence of the crown groups by 50 million years. These methods can be applied to date any previously characterised WGD event, including those identified in yeasts and vertebrates.

Chapter 4

Evolution of morphological disparity in the plant kingdom

James W. Clark, Alexander J. Hetherington, Jennifer L. Morris, Silvia Pressel, Jeff Duckett, Harald Schneider, Mark N. Puttick, Paul Kenrick, Dianne Edwards, Charles H. Wellman and Philip C.J. Donoghue

1 Summary

Patterns of phenotypic diversity (disparity) in the plant kingdom have fascinated biologists yet on the largest scales have not been explored. The distribution of disparity between taxa and the tempo of morphological evolution over time can answer fundamental questions about the evolution of plants. I present a disparity analysis of all major lineages of land plants based on a morphological dataset that covers all aspects of plant form. Subjected to multivariate ordination I show that the green plant morphospace is very 'clumpy' and major lineages of land plants are highly distinctive. This heterogeneity has partly resulted from extinction of fossil lineages, though some of the distances are maintained suggesting that large innovations divide the major land plant lineages. I reject a model that disparity peaks during the early evolution of the kingdom and show that continued innovations throughout the evolution of plants have resulted in continuous exploration of novel morphospace. I show how the phylogenetic signal is distributed among different sets of characters, and how some of these sets reveal divergent and convergent evolutionary trajectories.

The coding of the matrix was a collaboration between the author, JM and AJH, with advice and assistance from contributors. All analyses and the writing of this chapter were performed by the author.

2 Introduction

The study of disparity in the plant kingdom, and comparative morphology more broadly, can trace its theoretical roots to the work of Goethe (Kaplan 2001), who recognised homology between plant organs and characterised the 'Bauplan' that underlies the enormous diversity of observed forms (Mueller *et al.* 1952). Despite these early contributions, the evolution of plant disparity has remained overshadowed by the empirical observations derived from the study of animals (Hughes *et al.* 2013; Oyston *et al.* 2015; Oyston *et al.* 2016). Renewed interest in plant disparity has served to highlight the paucity of examples addressing the subject (Chartier *et al.* 2014; Oyston *et al.* 2016; Chartier *et al.* 2017), yet studying disparity has the potential to reveal how lineages adapt and change in response to external events as well as how internal factors limit or promote phenotypic change.

The current theoretical background for the evolution of disparity is derived principally from analyses of animal datasets. While general trends in macroevolution are rare, several patterns appear consistent across studies. Firstly, as with species diversity, disparity is not equally distributed between lineages. Secondly, lineages display a 'clumpiness', in that phenotypic variation is not a continuous and free mixing of traits (Raup and Gould 1974; Erwin 2007). Taxa tend to share certain morphologies, while other combinations of characteristics are never observed. Phylogenetic conservatism leads to closely related taxa sharing morphological characters while convergent evolution between distantly related lineages produces the opposite pattern. This clustering of taxa and the vacant spaces between has been explained in terms of constraint, extinction or a combination of the two (Erwin 2007).

The inclusion of fossil taxa in animal datasets facilitated vertical comparisons of disparity through time. As new species evolve and diverge towards the present, a traditional model predicts that disparity would increase concurrently. However, many animal clades show the opposite trend, with high levels of initial disparity that decrease towards the present (Hughes *et al.* 2013). Based on observations from the Burgess Shale lagerstatten, Gould (Gould 1990) sought to explain the 'clumpiness' of observed disparity through the

'diversification and decimation' model. Early radiations rapidly explore the limits of developmental possibility and produce maximal initial disparity, which is subsequently whittled away by the extinction of lineages. This results in a pattern of maximal initial disparity which decreases over time leaving a patchy extant morphospace.

Disparity can be represented by the distances between taxa in a morphospace, a mathematical space which is effectively reduced to 2 or 3 dimensions. Disparity is measured within a sample, and so is always relative to that sample, rather than being a property of any individual or species (Oyston *et al.* 2016). The shape of empirical morphospaces reflects this sampling and so they are dependent on both the characters and the taxa included (Wills 2001; Chartier *et al.* 2014; Mitteroecker and Huttegger 2015). Subsets of characters relating to specific functions, such as pollen or vasculature, or sampling within plant subclades allow a more practical and theoretically straightforward assessment of homology and geometric characters can be used. Sampling from a more diverse array of taxa becomes problematic as homology becomes increasingly difficult to establish - landmarks cannot be assigned when whole suites of characters are absent or not applicable (Hetherington *et al.* 2015). Despite this, new methodologies are pushing the capabilities of shape and outline analysis further to be able to sample specific characters such as leaf shape between phyla (Li *et al.* 2017). To quantify a morphology in a more holistic sense, cladistic characters can be used (Wills *et al.* 1994; Hetherington *et al.* 2015). Rather than describing shape, cladistic characters define homologies and can describe and differentiate states of a character as well as recording its presence or absence and applicability.

Disparity studies within plants have typically focussed on single aspects of morphology such as flowers, pollen, leaf shape and vasculature. Studies which have sought to quantify morphology in a more holistic manner have so far been restricted to subclades or individual lineages (Oyston *et al.* 2016). The general principles guiding the evolution of disparity are rooted almost entirely in studies of the animal kingdom. Plants represent an independent origin of multicellularity, and so an alternative system with which to compare the kingdom-level evolutionary dynamics of disparity. I present an analysis of the patterns of

morphological evolution across the entire plant kingdom, based on a dataset that is representative of a brange range of aspects of morphology. I show horizontal comparisons between extant clades, as well as vertical comparisons of disparity through time to test for a pattern of maximal initial disparity within the plant kingdom.

3 Materials and Methods

3.1 Matrix Assembly

An initial character matrix was assembled to span the Viridiplantae (Chlorophyta + Streptophyta) by fusing the character matrices from cladistic studies of green algae, charophytes and bryophytes (Mishler *et al.* 1994), tracheophytes and lycophytes (Kenrick and Crane 1997), ferns (Schneider *et al.* 2009), seed plants and gymnosperms (Hilton and Bateman 2006) and basal angiosperms (Doyle and Endress 2010). The characters represented a broad range of areas of plant morphology and tissue types (molecular and mitotic, sporophytic and gametophytic, biochemical, roots, shoots, sporangia, flowering, zoospores and sperm, spores, embryology and seeds). Overlapping characters were reconciled between matrices to avoid repetition and the number of character states expanded to capture morphology across a greater number of clades. Additional taxa and characters were added to the matrix and in many cases the scoring of taxa was revised in light of more recent understanding of homology or re-examination of taxa (Appendix 3). In total 56 new characters were implemented, 30 existing characters were combined between previous matrices in composite characters and 68 existing characters were modified (Appendix 3).

3.2 Inference of Missing Data

The distances between taxa were being poorly represented due to the non-random distribution of missing data in fossil taxa. I performed phylogenetic reconstruction under a Bayesian framework using the MKv+Γ model, in which the positions of extant taxa were constrained based on evidence from molecular systematics (Puttick *et al.* 2018), but the

placement of fossil taxa was unconstrained. I ran 4 parallel chains for 10 million generations each and selected 100 random trees from the posterior distribution. Along each tree, I simulated the possible tip states using stochastic character mapping (Huelsenbeck *et al.* 2003; Revell 2012). Stochastic character mapping calculates the conditional likelihood of each character state at each node in the tree, stochastically assigning node states based on their probability before simulating character history along each branch. I fixed known tip states with a probability of 1, and for unknown and missing tip states allowed each possible tip state an equal prior probability. I ran 1000 simulations per character per tree, and for each selected the state which had been sampled most frequently. I then estimated the most probable tip state and node state across all 100 trees in order to create a focal matrix which formed the basis for subsequent analyses.

3.3 Constructing the Viridiplantae morphospace

All ordination analyses were performed on the focal matrix. The distances between taxa were calculated using Gower's similarity metric (Gower 1971), which treats all character states as unordered and can tolerate missing data from the matrix. In addition, Gower's index does not count matching zeros in the calculation of dissimilarity, and so shared inapplicable characters do not contribute to the distance between taxa or their position within the morphospace (Deline 2009). The distance matrix was subjected to a Non-Metric Multidimensional Scaling (NMDS) multivariate analysis, with the number of dimensions constrained to 2. A stress plot was used to assess to extent to which the data was well represented in two dimensions and reported a strong relationship between the observed dissimilarity and the ordination distances ($r^2 = 0.99$). Non-metric methods are well suited for ordinations with a large proportion of absences and non-ordered multistate characters but produce a morphospace that can be challenging to interpret, as the resulting space is non-Euclidean and the distances between taxa are non-metric. I repeated the analysis using metric methods, using the Euclidean distance between taxa and a Principal Co-ordinates Analysis (PCoA), to test whether the NMDS analysis approximated a metric morphospace.

The morphospace was constructed initially with only extant taxa, and then with the inclusion of fossil taxa.

A consensus phylogeny based on molecular evidence and our current understanding of the placement of fossil taxa was used to construct a phylomorphospace (Kenrick and Crane 1997; Hilton and Bateman 2006; Doyle and Endress 2010; Revell 2012). The position of the nodes within the morphospace were based on distance between nodes and living taxa combined in a single ordination. Convex hulls were fitted around each major lineage to illustrate the occupied envelope of morphospace using vegan.

3.4 Disparity Between Lineages

Indices of disparity were calculated from the distance matrix. The disparity within lineages (mean disparity) was calculated as the mean pairwise distance between each taxon within the lineage. The partial disparity represents the contribution to the total morphological diversity of the kingdom and is calculated as the mean distance to the overall centroid for each taxon within a subclade, divided by N-1, where N is the total number of taxa included in the analysis (Foote 1993; Chartier *et al.* 2017). The sum of all partial disparities of all subclades is equal to the total morphological disparity of the entire sample. All calculations were performed on a sample of 1000 bootstrap replicates of the distance matrix and were performed using the *dispRity* package (Guillerme 2015).

3.5 Disparity Through Time

Calculation of disparity through time was performed using the time-slicing approach (Guillerme *et al.* 2018). I used a time-calibrated phylogeny containing 40 fossil taxa whose phylogenetic position could be robustly inferred. Analyses were based on the dissimilarity matrix and included the reconstructed ancestral node states for the phylogeny. I ran both punctuated and gradual models of evolution, with the punctuated model randomly selecting both accelerated and decelerated transformations. The matrix was bootstrapped 1000 times to estimate the standard error at each time point.

3.6 Dividing the morphospace

Characters within the matrix were subdivided in 8 broad and non-mutually exclusive categories: sporophytic (250 characters), gametophytic (56 characters), branching and appendages (55 characters), shoot anatomy (45 characters), roots and symbionts (20 characters), zoospores and spermatozoids (97 characters), spores, pollen and embryology (93 characters) and sporophylls and sporangia (58 characters). I recalculated a distance matrix for each subset of characters and produced an ordination using the same methods as outlined above. An initial morphospace produced for branching anatomy and appendages was heavily distorted by the lack of homology between euphyllophytes and other land plants, and so the space was recreated solely for euphyllophytes.

4 Results

I have utilised, revised and expanded character matrices that span the Viridiplantae kingdom and assembled a supermatrix that describes all aspects of their morphology (Mishler and Churchill 1985; Kenrick and Crane 1997; Hilton and Bateman 2006; Schneider *et al.* 2009; Doyle and Endress 2010). The matrix consisted of 548 binary and multistate characters to describe 248 living taxa sampled from each phylum resulting in 135,904 data points (Appendix 4). The vast diversity of angiosperms makes proportional sampling difficult, though our sampling approximately reflects known species diversity (Spearman's ρ = 0.83, p = 0.01). This matrix acts as the basis for calculating the dissimilarity between taxa, based on the shared absence and presence of character states. Subjected to non-metric multivariate ordination, the morphological disparity of the plant kingdom is presented in two dimensions (Fig 3.1). Non-metric morphospaces should be interpreted with caution as the distances between taxa are not Euclidean. I repeated analyses with a metric ordination method (Principal Components Analysis) and found that the distances between the taxa in both metric and non-metric ordinations were highly correlated (Mantel test, r = 0.99, p < 0.001). While the ordinated differences between taxa should still be interpreted qualitatively, the morphospace shows that the greatest distances between groups separate the land plants

(Embryophyta) from green algae, vascular and non-vascular plants, seed plants from spore plants and flowering plants from non-flowering plants.

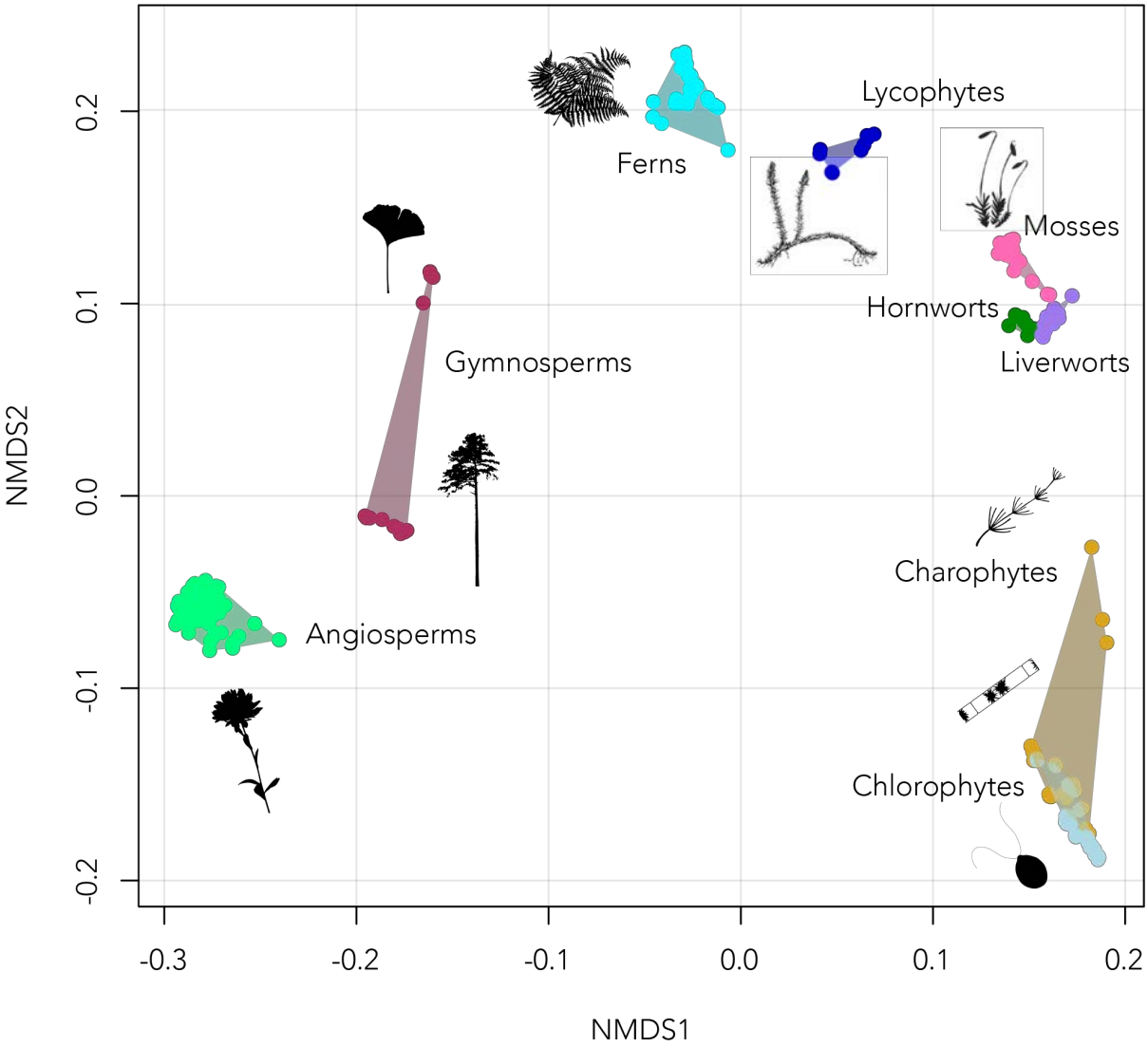


Figure 4.1. Empirical morphospace of the plant kingdom. The axes summarise morphological disparity derived from the observed dissimilarity between taxa (calculated using Gower's index) subjected to non-metric multidimensional scaling (NMDS). A convex hull was fit around each major lineage. Images of major lineages were obtained from PhyloPic.

The charophyte algae, here represented by non-embryophyte streptophytes, showed the highest mean disparity (Fig 4.2), although this is a paraphyletic grouping and much of the observed disparity is accentuated by the difference between the macrophytic and

unicellular taxa (Fig 4.1). The relatively low disparity among angiosperms compared to gymnosperms is surprising as, superficially, much of plant morphological diversity is considered within the context of floral characters. This in part reflects our sampling considering only the more basal lineages and that much of the disparity is contained within the more recently derived and diverse core eudicots and commelinid monocots. However, it is possible observed pattern may not be artefactual as the bodyplan of the angiosperms is relatively conserved (Floyd and Bowman 2007), at least within the context of the entirety of plant evolution.

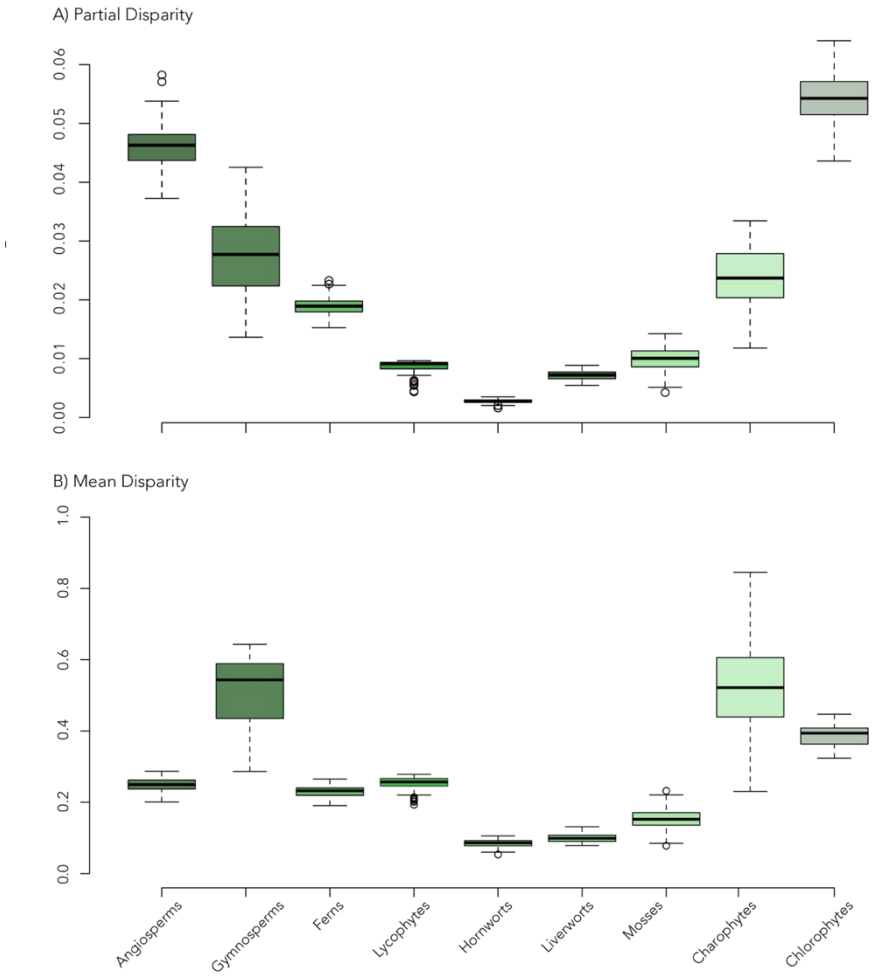


Figure 4.2. The partial disparity (A) represents the contribution of each lineage to the total morphological variation, calculated as the mean distance to the overall centroid standardised by the size of the subclade (Chartier *et al.* 2017). Mean disparity (B) is calculated as the mean pairwise distance between taxa within each lineage (Guillerme 2015).

High partial disparity can be explained by lineages possessing a relatively high mean disparity, or by lineages occupying extreme positions within the morphospace (Chartier *et al.* 2017). The two lineages with the highest partial disparity, the angiosperms and chlorophyte algae, both show low mean disparity, but are positioned at the extremes of the first axis (Fig 4.1). Indeed, when only plesiomorphic characters that are shared across all lineages are considered, the chlorophytes show the highest disparity, and the angiosperms among the lowest. Plotting the phylogenetic relationships between taxa onto the morphospace provides a means of exploring the phylogenetic distribution of morphological disparity (Fig 4.3). Our data show a strong phylogenetic signal, with the first axis dividing phyla in the order in which they have diverged (Puttick *et al.* 2018). There is scarcely any overlap between phyla, save for the green algal lineages, which are paraphyletic (Puttick *et al.* 2018).

A morphospace constructed with both fossils and extant taxa shows that no fossil taxa lie beyond the current envelope of morphospace and the fundamental shape of the morphospace is preserved. Within lineages, certain fossil taxa remain nested within the morphological envelope of their extant relatives, as seen in the fossil angiosperms (Fig 4.7). However, some fossil groups occupy intermediate positions between the clusters of extant taxa, such as the progymnosperms, pteridosperms and early tracheophytes.

Disparity through time analyses firstly reject a model of maximal initial disparity within the plant kingdom (Fig 4.8). Present levels of disparity are higher than at any point during plant evolution. Rates of morphological evolution are heterogeneous: an initial period of low variance is followed by a rapid increase throughout the Palaeozoic that likely coincides with three major evolutionary floras: the seedless plants during the Devonian (Palaeophytic), gymnosperms and early seed plants during the Mesozoic (Mesophytic) and the rise of angiosperms during the Jurassic/Cretaceous (Cenophytic) (Silvestro *et al.* 2015). Disparity continues to increase at a slower rate throughout the Devonian and Permian. A sharp increase during the Triassic is likely to represent the divergence of modern gymnosperms and the appearance of angiosperms. Disparity continues to increase towards the present, though again at a slower rate.

5 Discussion

The distinctiveness of each of the major lineages of land plants reflects their own distinctive developmental trajectories (Harrison 2017). The angiosperms contain the lowest proportion of non-applicable characters (34% across the sample), which suggests that the large number of innovations have led to the exploration of new areas of morphospace. By contrast the chlorophytes have the highest proportion of non-applicable characters (84%), and so it appears that the retention of a large number of plesiomorphic characters in the chlorophyte algae compared to the number of innovations leading to Streptophyta has led to them occupying a peripheral position.

Though sometimes correlated, species diversity and morphological disparity are often decoupled (Wagner 2010; Minelli 2016; Romano *et al.* 2017). It has been proposed that a robust or constrained bodyplan can lead to increased diversification (Rabosky *et al.* 2012; Melzer and Theissen 2016), which would predict a negative relationship between lineage diversity and disparity. I report that within the plant kingdom, disparity is not correlated with species diversity (Spearman's $\rho = 0.23$, $p = 0.55$), though certain clades, such as the hornworts, do show low levels of both disparity and diversity (Villarreal *et al.* 2014), and it is likely that a more comprehensive survey of angiosperms would only increase their apparent disparity. Further to this, I found no correlation between the absolute age of each lineage and the mean disparity (Pearson's $r = 0.541$, $p = 0.16$), though the phylogenetic (patristic) distance between taxa was correlated with their morphological distance (Mantel test, $r = 0.3$, $p = 0.001$). Despite this, there are instances where convergence is more apparent than conservatism such as the position of *Chara* as the algae most morphologically similar to the embryophytes (Fig 4.3).

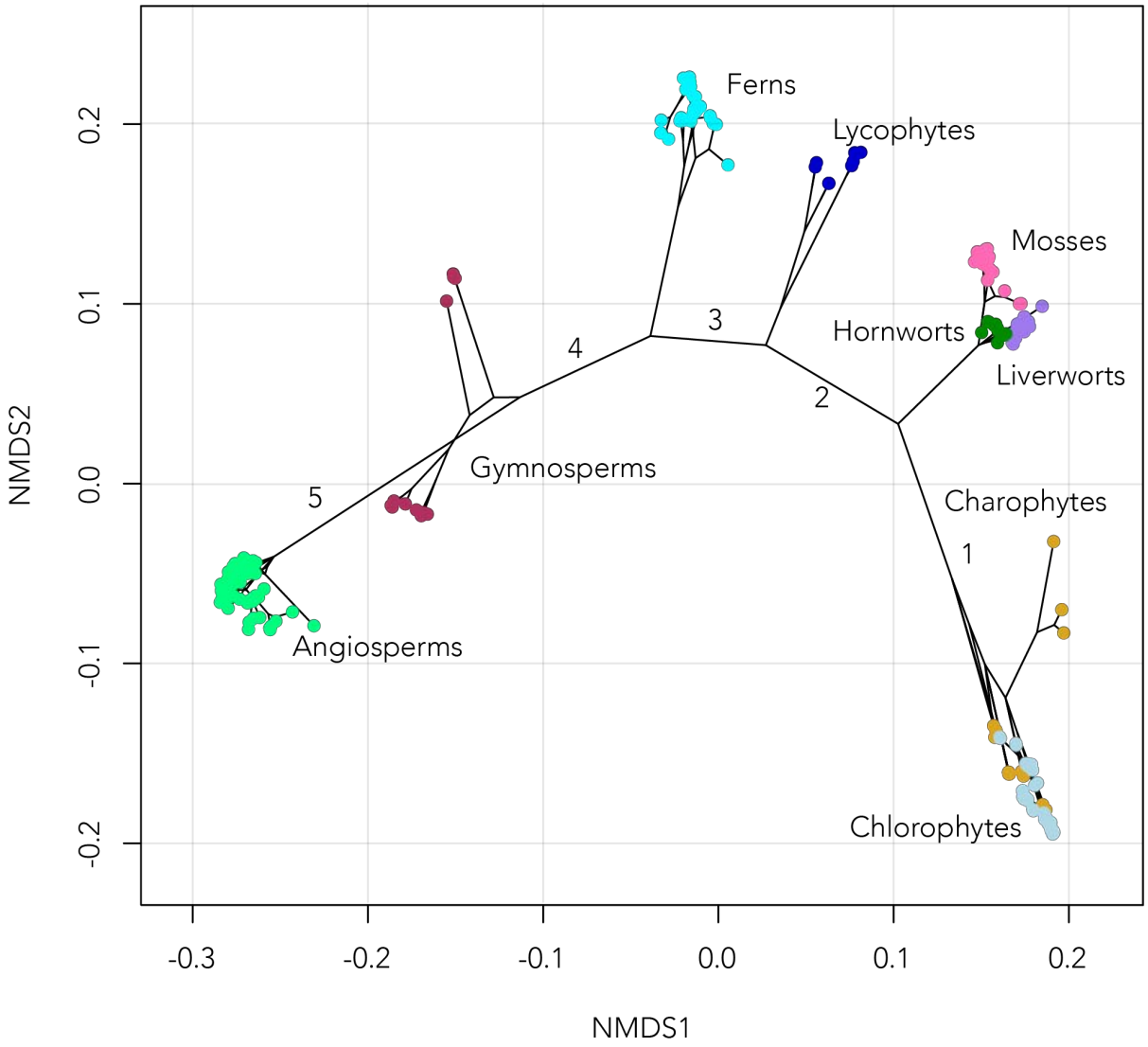


Figure 4.3. A phylomorphospace of extant land plant diversity. The tree represents a summary of the current hypotheses of phylogenetic relationships, and the character states at each node were estimated through stochastic character mapping across a sample of trees. Numbers represent 1) The movement to land (Embryophytes), 2) The origin of vascular tissue (Tracheophytes), 3) The origin of 'true' leaves (Euphyllophytes), 4) The origin of seeds (Spermatophytes) and 5) The origin of flowers (Angiosperms).

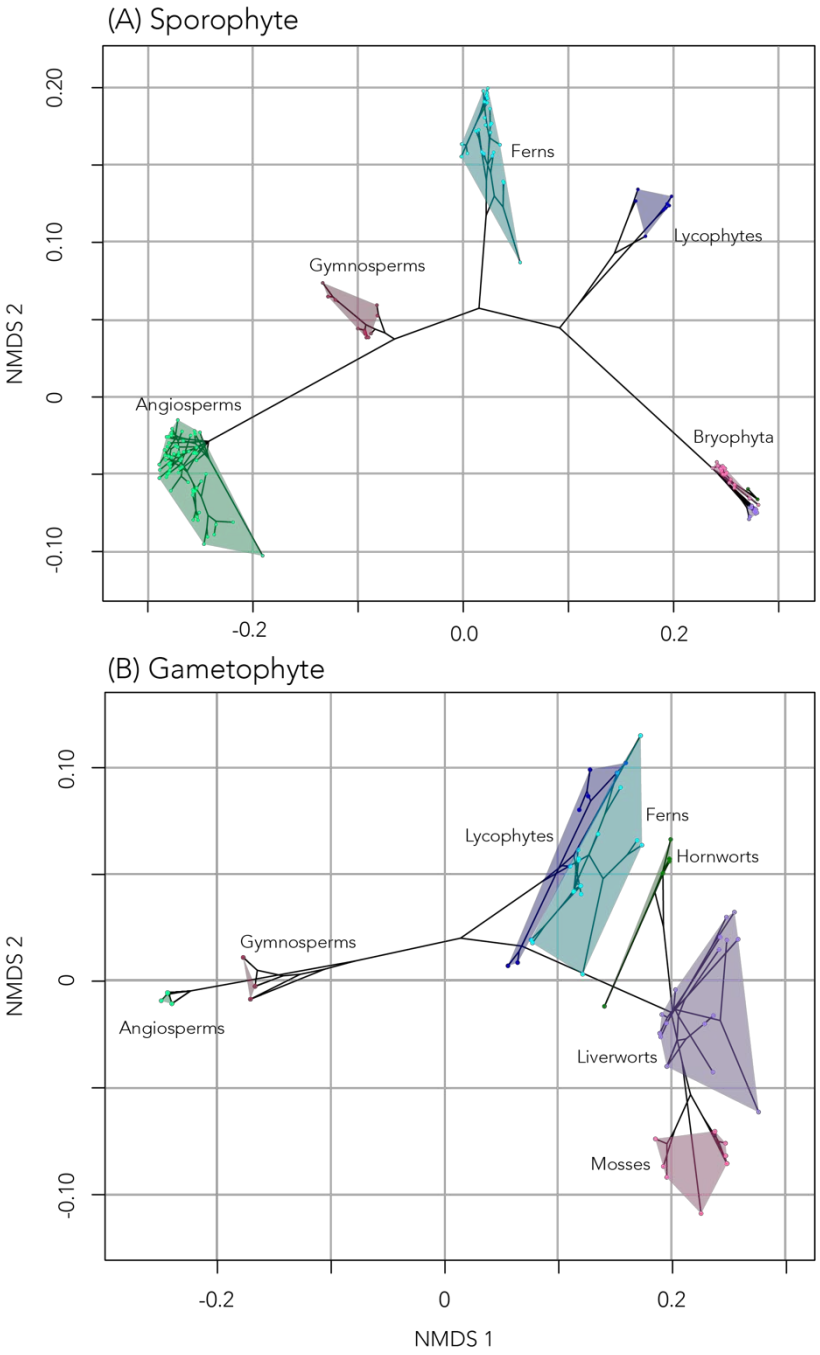


Figure 4.4. A comparison of the gametophyte and sporophyte morphospace in land plants (Embryophyta). The axes summarise morphological derived from the observed dissimilarity between taxa (calculated using Gower's index) subjected to non-metric multidimensional scaling (NMDS). Sets of characters relating specifically to (A) the sporophyte and (B) the gametophyte characters were subsampled from the morphological matrix. A convex hull was fit around each major embryophyte lineage.

Morphospaces constructed around single characteristics, such as leaf shape or flowers, have demonstrated much greater convergence between phyla, though still with some phylogenetic signal (Boyce and Knoll 2002; Chartier *et al.* 2014; Li *et al.* 2017). Dividing our matrix into different suites of characters produced very distinctive morphospaces with contrasting patterns. I initially divided the morphospace to reflect the two life cycles of land plants. The alternation between multicellular diploid and haploid phases is a defining trait of land plants (Doyle 2012). The earliest diverging lineages all possess dominant haploid

gametophyte phases, with sporophytes incapable of independent growth (Doyle 2012). More recently diverged lineages have progressively decreased the size and dominance of the gametophyte, while simultaneously increased the independence of the sporophyte. Deep homologies exist between the two, and the sporophyte phase is thought to have elaborated as a heterochronic change by delaying the onset of reproductive stages and has co-opted developmental pathways present in an ancestral gametophyte (Pires and Dolan 2012; Tomescu *et al.* 2014). The evolution of a branched sporophyte is believed to have facilitated further innovation within the vascular plant lineage (Tomescu *et al.* 2014). By subsampling the morphological matrix, the relative disparity of the two stages can be compared between lineages. Intuitively, bryophytes possess the most diverse gametophytes but have the lowest sporophyte disparity (Fig 4.4). A contrasting pattern is observed in the seed plants, whose highly reduced gametophytes result in them occupying the smallest, yet still highly distinct, regions of morphospace. The lycophytes and ferns show broad morphospace occupation in both generations: the fern sporophyte is more similar to that of the seed plants, but the gametophyte is closer to the bryophytes. Conversely, the lycophyte sporophyte is more similar to the bryophytes while the gametophyte generation bears more similarity to the seed plants. Overall, the patterns of morphospace occupation mirror the evolutionary and developmental trajectories: through time the sporophyte has become increasingly elaborate seemingly at the expense of the gametophyte.

Interestingly, morphospaces built around vegetative characters (stem anatomy, branching and appendages) show a much less clear phylogenetic signal and greater convergence between lineages and divergent evolution within lineages is more apparent (Fig 4.5). The reproductive character sets tend to reinforce the distances between lineages (Fig 4.6), with these distances most exaggerated between the pollen, spore and embryology characters. These results reflect that the degree of convergence between lineages is much greater in vegetative traits, and the same morphologies have likely evolved repeatedly in response to extrinsic pressures. The reproductive characters appear to be driving the large

distances between the clades and indicate that many of the key innovations that distinguish land plant lineages have been reproductive.

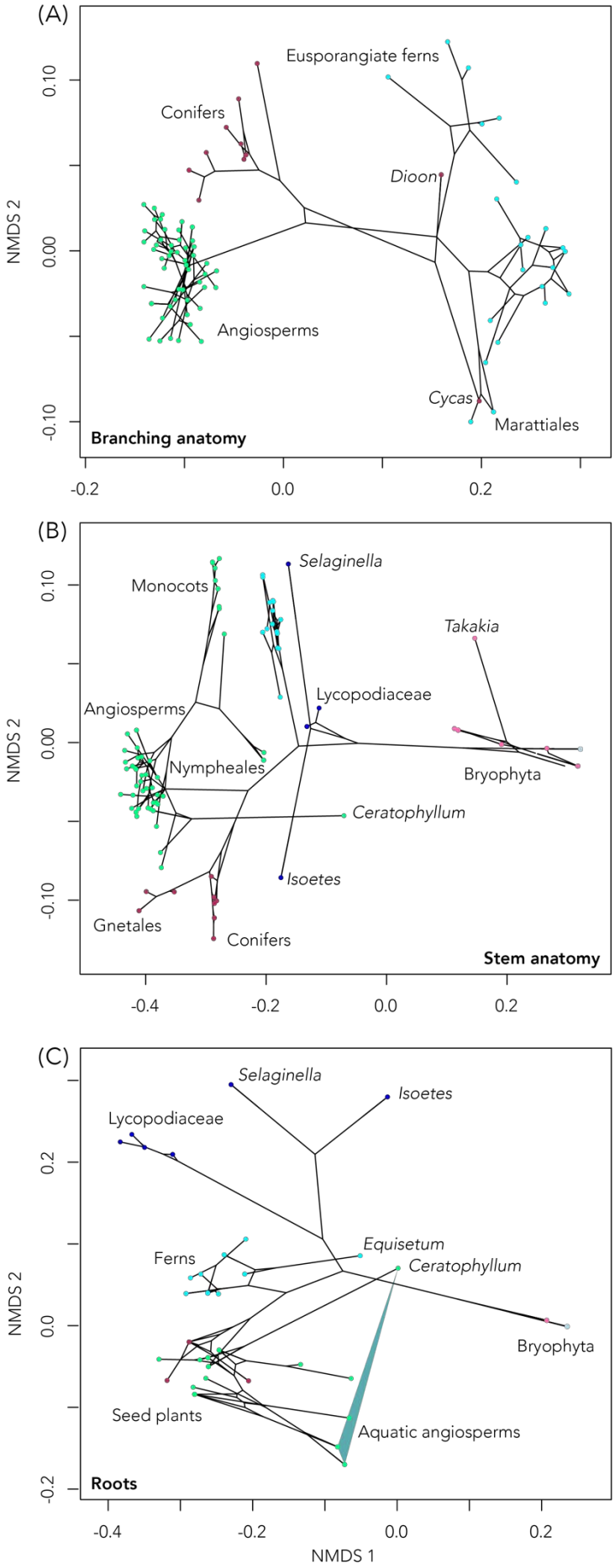


Figure 4.5. Vegetative morphospaces. Subsets of characters from the original matrix were sampled to produce morphospaces for (A) branching anatomy among euphyllophytes, (B) stem anatomy and (C) roots and endosymbionts. Notable taxa have been labelled and certain groupings have been highlighted under convex hulls.

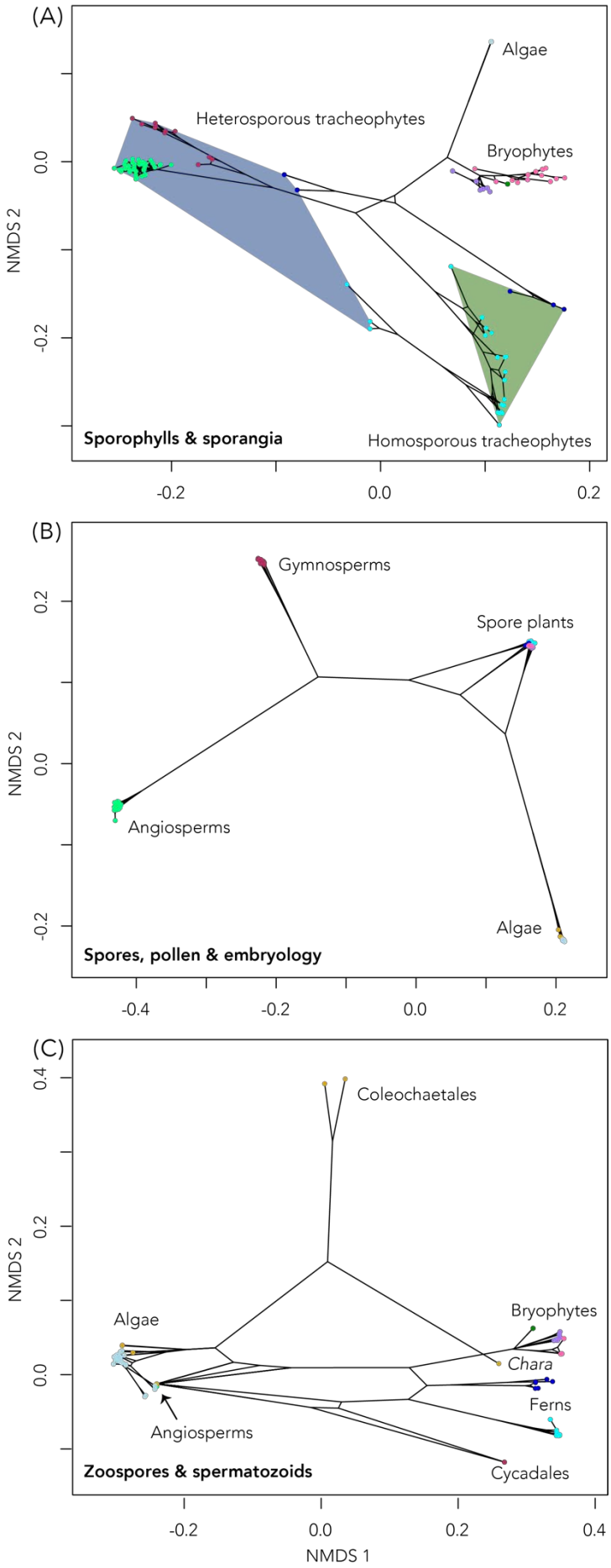


Figure 4.6. Reproductive morphospaces. Subsets of characters from the original matrix were sampled to produce morphospaces for (A) Sporophylls and sporangia (B) Spores, pollen and embryology (C) zoospores and spermatozoids. In (A) convex hulls were fitted to highlight the distinction between homosporous and heterosporous tracheophytes. Notable taxa have been labelled and certain groupings have been highlighted under convex hulls.

The fossil record contains evidence of many unique character combinations that are not found among extant plants (Chomicki *et al.* 2017). The large inter-phylum distances in the extant morphospace (Fig 4.1) may represent impossible designs, though an alternative hypothesis is that they are artefacts of extinction and that fossil taxa may occupy the intermediate spaces between crown groups. I introduced 160 fossil taxa including cryptophytes, early tracheophytes, zosterophylls, lycopsids, progymnosperms and pteridosperms (seed ferns), as well as fossils that are assigned to the major extant lineages. The inclusion of fossil taxa introduces further challenges as plant macrofossils are rarely informative about the entire plant, resulting in large proportions of non-random missing data. Dissimilarity indices, such as Gower's coefficient, can accommodate missing data to some degree (Deline 2009), yet our data showed that fossil taxa clustered in distinct areas of morphospace. Phylogenetic inference of missing data allowed a conservative estimate of probable character states based on a current hypothesis of land plant relationships.

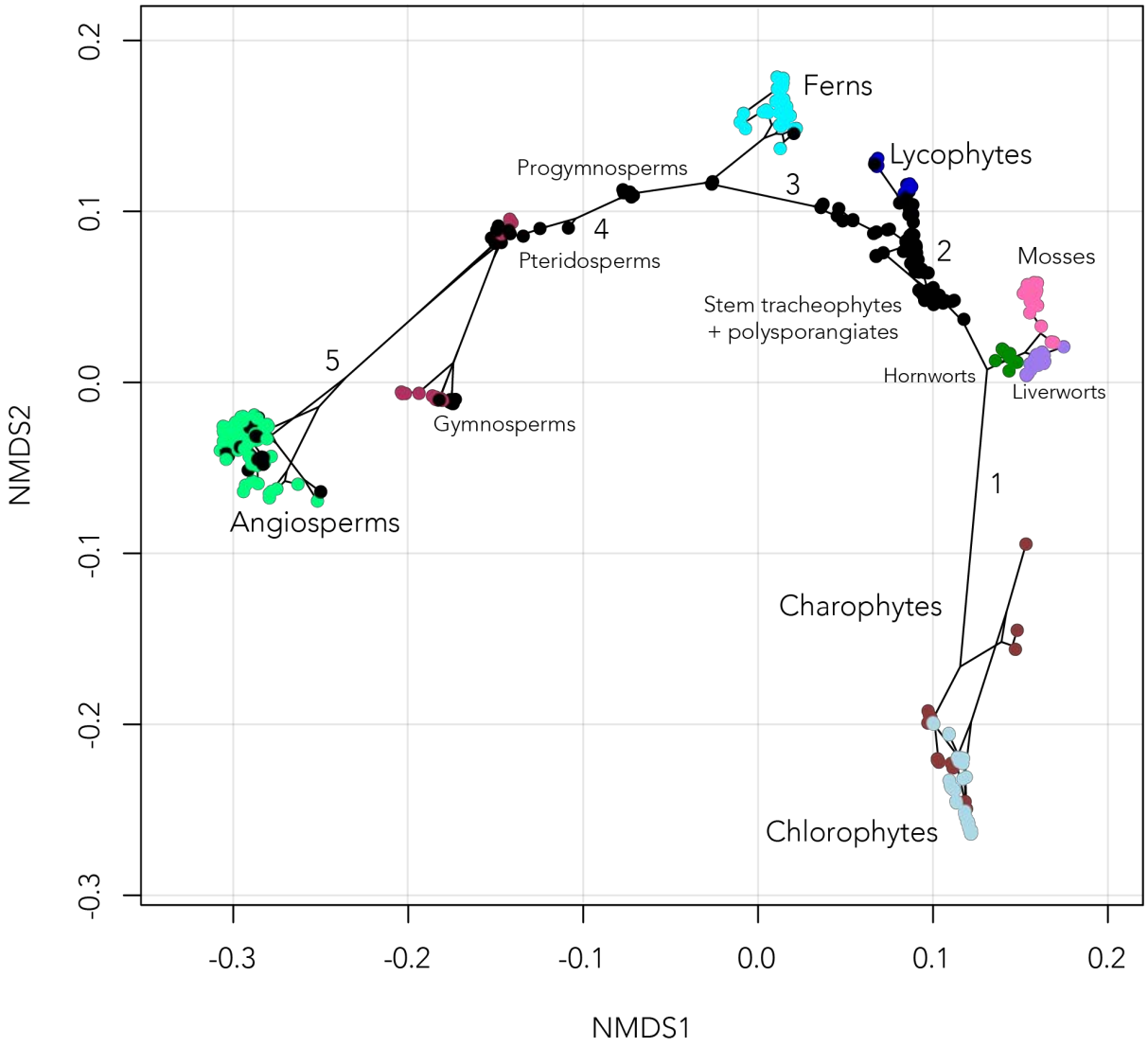


Figure 4.7. Empirical phylomorphospace of the plant kingdom included fossil taxa. The axes summarise morphological disparity derived from the observed dissimilarity between taxa (calculated using Gower's index) subjected to non-metric multidimensional scaling (NMDS). A convex hull was fit around each major lineage. Fossil taxa are shown as black dots. The tree represents a summary of the current hypotheses of phylogenetic relationships, and the character states at each node were estimated through stochastic character mapping across a sample of trees. Numbers represent 1) The movement to land (Embryophytes), 2) The origin of vascular tissue (Tracheophytes), 3) The origin of 'true' leaves (Euphyllophytes), 4) The origin of seeds (Spermatophytes) and 5) The origin of flowers (Angiosperms).

The position of these extinct lineages supports the hypothesis that some of the morphological innovations in land plants have been sequential (Mishler and Churchill 1985). As predicted by hypotheses of sporophyte evolution (Tomescu *et al.* 2014), early tracheophytes and polysporangiates are morphologically intermediate between modern bryophytes and tracheophytes (Fig 4.7). The nature of both generations of the earliest polysporangiates is uncertain. While they possess branched sporophytes, many appear too small to be independent of the gametophyte (Boyce 2008). More recent discoveries have highlighted putative polysporangiates that may have been large enough to exist independently of the gametophyte (Libertin *et al.* 2018). These early polysporangiates have been predicted to occupy an intermediate morphology, and indeed our results place them between vascular and non-vascular plants (Fig 4.7). Other morphological distances, such as that between the angiosperms and gymnosperms, are maintained even with the inclusion of fossils and support a scenario of 'saltational' or punctuated evolution. This distance represents the large number of innovations that characterise angiosperms: the evolution of the flower and a large number of vegetative innovations (Doyle 2012), but also the absence of any intermediate morphologies present in fossils assigned unequivocally to the angiosperm stem lineage (Doyle 2008; Herendeen *et al.* 2017).

Maximal initial disparity predicts that the limits of morphospace are realised early during the evolution of the kingdom, and that subsequent lineages have contributed little, if anything, to the total disparity. Certain elements of the plant developmental toolkit show a high level of conservation (Floyd and Bowman 2007), which suggests that the potential for generating high morphological disparity is present in some the earliest diverging lineages of land plants. However, some developmental regulators have evolved in a highly lineage-specific pattern and have led to some of the more recent innovations in land plants (Chanderbali *et al.* 2016). Studies based on extant taxa have reported that maximal disparity was attained early during the evolution of angiosperms, conifers and ferns and has remained largely independent of species diversification (Oyston *et al.* 2016). The earliest records of land plants are less striking than the metazoan Cambrian explosion, yet the Silurian-

Devonian diversification of vascular plants was accompanied by the appearance of several unique bodyplans (Bateman *et al.* 1998; Chomicki *et al.* 2017).

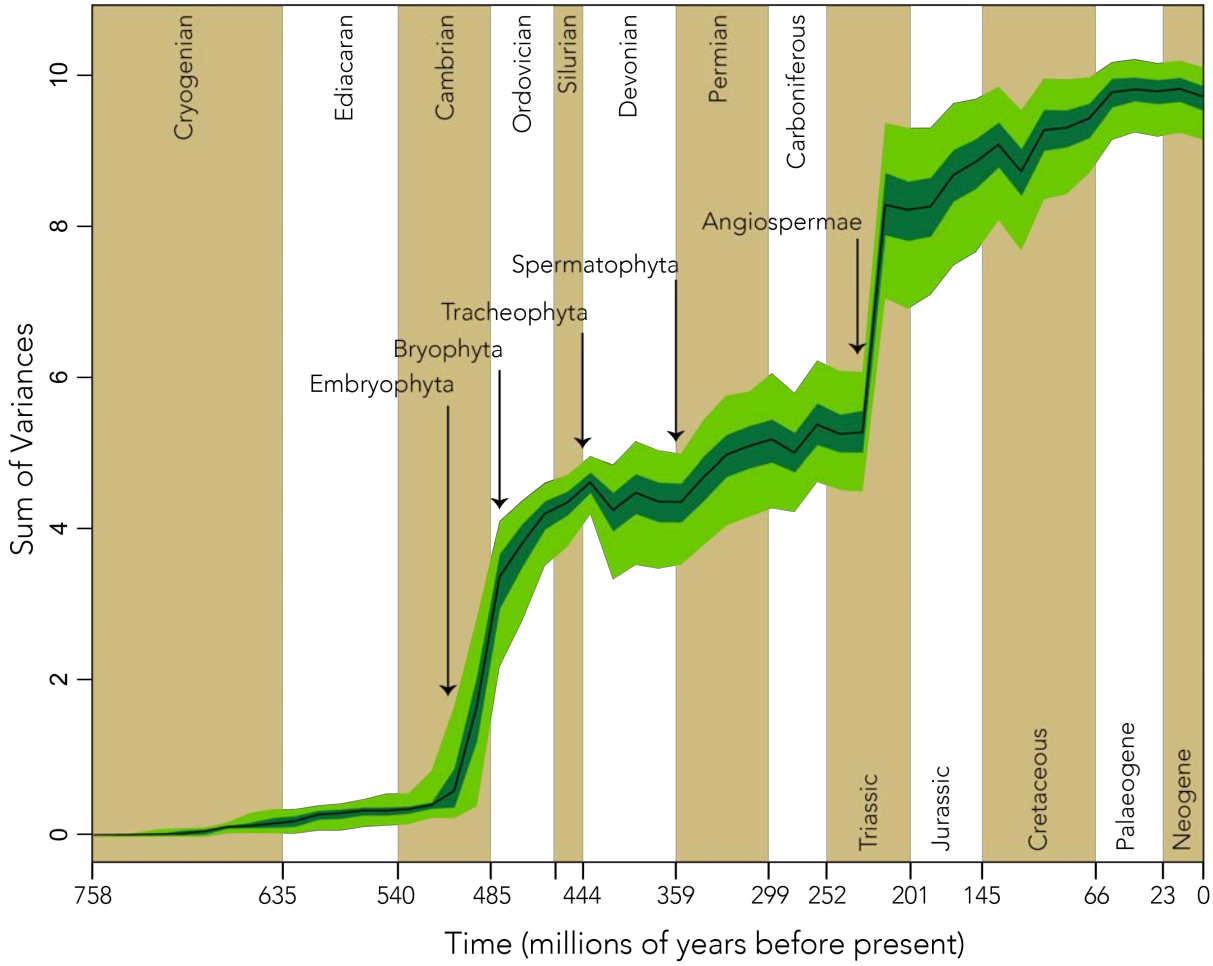


Figure 4.8. Disparity through time. The sum of variances represents the trajectory of disparity through geological time, estimated using a time-calibrated phylogeny including fossil taxa whose phylogenetic position could be reliably estimated. Species divergence dates were obtained from Morris *et al.* (2018). Disparity through time was estimated using *dispRity* (Guillerme 2015) in R, under a model of gradual evolution with time bins evert 50 million years. Arrows show the approximate timing of the origin of several major clades of land plants.

Disparity through time indicates both periods of rapid morphological innovation and relative stasis, separated by the arrival of the evolutionary floras (Fig 4.8). These major floras are characterised by a succession of key innovations: true leaves, the seed and the flower respectively. These successive innovations mean that unlike animals, disparity among plants has continued to increase throughout geological time, reflecting the pattern of ordinal diversity (Valentine *et al.* 1991).

Within certain lineages, including the angiosperms and gymnosperms, fossils are located at or close to the limits of the envelope (Fig 4.7). This suggests that high early disparity may be a pattern within lineages, followed by a plateau where space is 'packed' rather than expanded. The plateaus within clades could however be an artefact of using discrete characters: the number of novel character states can saturate and the coding of synapomorphies means that the number of unique character states in extant taxa is often underestimated. While extinction acts to reduce the continuity of morphological diversity, initial maximal disparity has also been explained in terms of constraints which may prevent the occupation of novel regions of morphospace. Constraints have been classified as geometric, ontogenetic, physical and ecological (Oyston *et al.* 2015). These classifications represent intrinsic (geometric and ontogenetic) and external (physical and ecological) factors that shape the evolution of a lineage but are difficult to disentangle. If maximal disparity was attained early during the evolution of each phylum, then at the kingdom level it appears that constraints can be overcome through the evolution of major innovations that have led to the colonisation of whole new regions of morphospace. Laboratory experiments have revealed that mutations can produce phenotypes consistent with some of the major transitions in land plants, including inducing multicellularity in green algae or branching in bryophyte sporophytes (Vivancos *et al.* 2012; Hanschen *et al.* 2016). It has also been suggested that major genomic events, including whole genome duplication (WGD) may provide the raw genetic material to facilitate the evolution of such large morphological jumps, including the evolution of flower (Chanderbali *et al.* 2016). Such experiments and events provide a means to understanding how some lineages of land plants escaped developmental constraint.

6 Conclusions

Patterns of morphological evolution within the plant kingdom can be quantified and analysed through the use of discrete characters. The resulting morphospace shows that plants have diversified and explored morphospace in a fundamentally different way to animals. In animals large areas of empty morphospace are the result of extinction, and in certain instances this holds true of plants also, yet some of the large distances between lineages in plants are the result of key innovations that have allowed a transformation of the bodyplan. Across the land plant phylogeny, different sets of characters reveal different patterns of divergence and convergence. Across lineages, the sporophyte and gametophyte life cycles show opposite trends in terms of the relative disparity. Vegetative characters tend to show convergence between lineages while reproductive characters emphasise the distances between taxa. Importantly, the inclusion of fossil taxa has revealed that the distances between some lineages can be explained by the extinction of intermediate forms. However, in the case of the angiosperms the distances are maintained, revealing that the lineage is highly distinctive. Finally, at the kingdom level I reject the hypothesis of maximal initial disparity and show that novelty has continued to evolve throughout plant evolutionary history.

Chapter 5

A History of Whole Genome Duplication in Poales

James W. Clark, Tom A. Williams and Philip C. J. Donoghue

1 Summary

Poales represent one of the most species rich and ecologically successful lineages of flowering plants. Their importance as agricultural crops and architects of ecosystems has led to a detailed narrative of extrinsic factors that have patterned their diversification. They also show an extensive history of whole genome duplication (WGD), with extant grasses having undergone three WGD events (*rho*, *sigma* and *tau*), as well as evidence for continued polyploidy and hybridisation amongst living species. The role of WGD in promoting diversification is poorly understood, yet, due to a comparably dense sampling of genomic data, Poales is one of the best opportunities to test for an effect of recurrent WGD on species diversity. I present a novel application of the Amalgamated Likelihood Estimation (ALE) method to detect and characterise WGD in a transcriptomic dataset across Poales, lending further support for multiple WGD events throughout monocots. I use evidence from the fossil record to estimate the ages of the *rho* WGD event (80-77 Ma), the *sigma* WGD event (136-126 Ma) and the *tau* WGD event (152-139 Ma). The nature of these WGD events in terms of single or multiple parent origins remains unclear. I perform a diversification analysis of over 7,000 species to show that despite multiple shifts in the rate of diversification across the monocot phylogeny, there is no strong signal of increased diversification in the wake of WGD. These results provide a timeline of genomic events throughout the history of Poales and question the role of WGD in promoting species diversity.

This chapter is unpublished. The experiments were designed by the author, TW and PD and performed, analysed and this written up by the author

2 Introduction

2.1 The evolution of Poales

Poales includes some of the most diverse families of flowering plants and are an order of unique economic importance. Comprising c. 22,000 species, they contribute approximately 7% of total plant diversity (Givnish *et al.* 2010). Species diversity within the order is unevenly distributed across 16 families, with vast majority concentrated within just two families: Poaceae (grasses) and Cyperaceae (sedges). The order comprises some of the most economically important and ecologically successful lineages of land plants. The grass family alone accounts for 60% of human energy intake and is estimated to cover 31-41% of the Earth's surface (Beer *et al.* 2010; Linder *et al.* 2018). Poales has a truly pan-global range and is found in almost every non-marine habitat, including deserts and Antarctica. Accompanying their rise to ecological success and ubiquity, Poales boasts a diversity of photosynthetic mechanisms with both C4 and CAM photosynthesis have evolved in multiple independent lineages (Edwards *et al.* 2010; Bouchenak-Khelladi *et al.* 2014).

The extrinsic drivers of diversification within Poales have been well documented (Humphreys and Linder 2013; Bouchenak-Khelladi *et al.* 2014; Linder *et al.* 2018). Our determination to understand the rise of grassland ecosystems stems from their role in the evolution of our own species (Bonnefille 2010). The diversification of the most species rich lineages has been correlated with the evolution of wind pollination and C4 photosynthesis (Poaceae and Cyperaceae) and with the evolution of an epiphytic life history and CAM photosynthesis in the Bromeliaceae (Bouchenak-Khelladi *et al.* 2014). In the case of C4 photosynthesis, this innovation coincides with periods of change in the Earth's atmosphere, with falling CO_2 levels triggering a rise of lineages with carbon concentrating mechanisms (Christin *et al.* 2008). Other radiations have been ascribed to the evolution of increased

tolerance to a range of climatic conditions, including cold tolerance (Humphreys and Linder 2013) as well as to fire (Bond *et al.* 2003).

2.2 A role for genome duplication

Whole genome duplication (WGD) has been posited as an intrinsic driver of diversification and macroevolution (Ohno 1970; Soltis and Soltis 2016). All extant members of the most recently derived family of Poales, Poaceae, have undergone at least 5 rounds of WGD (*zeta, epsilon, tau, sigma* and *rho*) and in addition many lineages have undergone subsequent WGD or genome triplication events (Jiao *et al.* 2011; D'Hont *et al.* 2012; Estep *et al.* 2014; Jiao *et al.* 2014; McKain *et al.* 2016). All non-Alismatoid monocots (Petrosaviidae) underwent one round of WGD (*tau*) (Jiao *et al.* 2011). A second event, *sigma*, is shared by all Poales, while all grasses (Poaceae) underwent a third (*rho*). Many important crop species have also undergone more recent WGD events that are associated with the traits amenable to domestication (Renny-Byfield and Wendel 2014; Salman-Minkov *et al.* 2016). A comparable history of WGD is found in other model lineages of flowering plants, such as Brassicales, where it has been shown that repeated rounds of WGD have contributed to the evolution of novel traits (Kagale *et al.* 2014; Edger *et al.* 2015).

The association between WGD and diversification has remained contentious: among extant species there is a trend for polyploids to exhibit higher rates of extinction (Mayrose *et al.* 2015). However, based on the occurrence of WGD events during the evolution of several hyper-diverse lineages, including Poaceae, it has been proposed that WGD may be important for promoting diversity (Soltis *et al.* 2014b; Mayrose *et al.* 2015; Tank *et al.* 2015). Initial tests for a direct association between WGD and increased rates of diversification found little support for this hypothesis, though it was observed that a short lag or latency period separates several WGD events from bursts of diversification (Schranz *et al.* 2012; Tank *et al.* 2015). These authors coined the 'WGD lag-time' model which proposed that WGD provides the intrinsic means to promote diversity and evolvability. Genome evolution post-WGD is characterised by gene loss, transposable element proliferation and rearrangements

(Dodsworth *et al.* 2016). The WGD-lag time model proposes that WGD may facilitate diversification during this period of diploidisation, with the latent period representing the time taken to diploidise (Schranz *et al.* 2012). Alternatively, it was also suggested that WGD may provide an increased evolvability that in suitable environmental context may promote increased success and diversification (Schranz *et al.* 2012).

An analysis of more recent WGD events and diversification within the grasses showed that over a relatively short period of 20 million years (myrs), extensive allopolyploidy (polyploidy that occurs alongside hybridisation) had no impact on the diversity of the grass tribe Andropogoneae (Estep *et al.* 2014). Furthermore, a recent survey across angiosperms found that fewer than half of known WGD events are associated with a shift in diversification (Landis *et al.* 2018). Together, these results would a tenuous role for WGD in the diversification of Poaceae, yet the presence of WGD at the base and throughout the evolution of one of the most successful lineages of flowering plants remains intriguing.

2.3 Characterising WGD

Clusters of genes within the genome can be mapped between species to determine syntenic blocks, and these methods robustly inferred the presence of three rounds of WGD in the history of grasses (Tang *et al.* 2010; Jiao *et al.* 2014). Phylogenomic methods look for a topological signal of WGD in the evolutionary history of gene trees (Jiao *et al.* 2011). Genes will undergo divergence and extinction in a pattern mirroring the species tree but also processes independent of the overlying phylogeny including loss, duplication and transfers that lead to a complex and sometimes reticulate history that deviates from the species tree (Galtier and Daubin 2008). Transfers represent horizontal gene transfers, achieved through microbial transmission or introgression. The combined evolutionary history of gene trees across species, or 'phylomes', highlights nodes where the number of duplication events is higher than a background rate (McKain *et al.* 2016), representing putative WGD events. Reconstructing gene family history is often undermined by the relative paucity of phylogenetic signal present in the sequence data, especially when long branches or periods

of rapid speciation are involved. Combined, these factors present a real challenge in reliably inferring the history of WGD using phylogenomic methods.

The model of reconciliation needs to account for the complex nature of gene family evolution, including the possibility that gene transfers could arise through hybridisation (A duplicate-transfer-loss model) (Stolzer *et al.* 2012). Allopolyploidy has been known to feature in the evolution of some grass lineages (Estep *et al.* 2014). To date, few characterised WGD events have been distinguished based on their nature; though novel methods now exist which determined the nature of even relatively ancient WGD events (Marcet-Houben and Gabaldon 2015; Gregg *et al.* 2017). The model also needs to account for topological uncertainty in the underlying gene trees (Bansal *et al.* 2018) – this has previously been approach *a priori* by filtering gene trees with only high support values (Jiao *et al.* 2011). Trees built from single genes are capable of producing erroneous topologies due to a lack of signal or biological phenomena such as incomplete lineage sorting. Further, due to the high volume of data involved in phylogenomic analyses, reconstructing gene family evolution relies on fast and heuristic methods, at the expense of accuracy.

WGD has played a major role in the evolution and domestication of our major food crops, yet the effects of deeper events remain unknown. Exploring hypotheses of macroevolutionary causality require that each event be well constrained phylogenetically and geologically. Here, I present a comprehensive analysis of genome evolution of Poales through time. I demonstrate a refined approach for locating WGD events within a phylogeny and estimate the absolute timing of three major WGD events during the evolution of grasses. Further, I explicitly test for an association between all three events and an increase in the rate of diversification.

Table 5.1. Full list of taxa with transcriptomes included in the study alongside their reference source. Transcriptomes from McKain *et al.* (2016) were used for reconciliation.

Taxon	Source	Order	Taxon	Source	Order	Taxon	Source	Order
Amborella trichopoda	McKain *et al.*	Amborellales	*Typha angustifolia*	McKain *et al.*	Poales	*Oropetium thomaeum*	PLAZA 4.0	Poales
Nuphar advena	1KP	Nymphealeas	*Brocchinia reducta*	McKain *et al.*	Poales	*Brachypodium distachyon*	McKain *et al.*	Poales
Kadusra heteroclita	1KP	Austrobaileyales	*Neoregalia carolinae*	McKain *et al.*	Poales	*Hordeum vulgare*	PLAZA 4.0	Poales
Canella winterana	1KP	Canellales	*Ananas comosus*	PLAZA 4.0	Poales			
Sarcandra glabra	1KP	Chloranthales	*Stegolepis ferruginea*	McKain *et al.*	Poales			
Vitis vinifera	McKain *et al.*	Vitales	*Juncus effusus*	McKain *et al.*	Poales			
Acorus americanus	1KP	Acorales	*Juncus inflexus*	McKain *et al.*	Poales			
Spirodela polyrhiza	PLAZA 4.0	Alismatales	*Mapania palustris*	McKain *et al.*	Poales			
Pistia stratiotes	1KP	Alismatales	*Lepidosperma gibsonii*	McKain *et al.*	Poales			
Triglochin maritima	1KP	Alismatales	*Cyperus alternifolia*	McKain *et al.*	Poales			
Posidonia australis	1KP	Alismatales	*Cyperus papyrus*	McKain *et al.*	Poales			
Zostera marina	1KP	Alismatales	*Mayaca fluviatilis*	McKain *et al.*	Poales			
Dioscorea villosa	1KP	Dioscoreales	*Lachnocaulon anceps*	McKain *et al.*	Poales			
Talbotia elegans	1KP	Pandanales	*Xyris jupicai*	McKain *et al.*	Poales			
Ludovia sp.	1KP	Pandanales	*Chondropetalum tectorum*	McKain *et al.*	Poales			
Freycinetia multiflora	1KP	Pandanales	*Elegia fenestrata*	McKain *et al.*	Poales			
Colchicum autumnale	1KP	Liliales	*Aphelia sp.*	McKain *et al.*	Poales			
Smilax bona-nox	1KP	Liliales	*Centrolepis monogyna*	McKain *et al.*	Poales			
Phalaenopsis equestris	PLAZA 4.0	Asparagales	*Flagellaria indica*	McKain *et al.*	Poales			
Hosta venusta	McKain *et al.*	Asparagales	*Ecdeiocolea monostachya*	McKain *et al.*	Poales			
Yucca filamentosa	McKain *et al.*	Asparagales	*Joinvillea ascendens*	McKain *et al.*	Poales			
Elaeis guinnensis	McKain *et al.*	Arecales	*Streptochaeta angustifolia*	McKain *et al.*	Poales			
Phoenix dactylifera	McKain *et al.*	Arecales	*Oryza sativa*	McKain *et al.*	Poales			
Sabal burmudana	1KP	Arecales	*Aristida stricta*	McKain *et al.*	Poales			
Musa acuminata	McKain *et al.*	Zingiberales	*Sorghum bicolor*	McKain *et al.*	Poales			
Zingiber oficinalis	McKain *et al.*	Zingiberales	*Dendrocalamus latiflorus*	McKain *et al.*	Poales			
Typha latifolia	McKain *et al.*	Poales	*Phyllostachys edulis*	PLAZA 4.0	Poales			

3 Materials and Methods

3.1 Divergence time estimates in Poales

A full phylogenomic dataset for Poales was presented by McKain *et al.* (2016), which included 33 monocot species, 27 belonging to Poales, and two outgroup species (*Vitis* and *Amborella*; Table 5.1). 222 single copy orthologues identified by McKain *et al.* were used as the input data for a molecular clock analysis. Fossil node calibrations were selected from the literature (Table 5.2), which were defined and justified following best practice (Parham *et al.* 2012; Iles *et al.* 2015). The topology was constrained to that inferred by McKain *et al.* (2018) (Fig 5.1). Minimum node age calibrations were modelled as a diffuse truncated Cauchy distribution, with a 1% probability tail that node ages may lie younger than the specified fossil. Calibrations that specify both a maximum and minimum node age were modelled as a uniform distribution with a soft maximum, a 2.5% probability tail that the node age can exceed the maximum age. The molecular clock analysis was performed using the approximate likelihood function in MCMCTREE and a GTR+Γ model (Yang 2007). A relaxed clock model was used, which allows the clock rate to vary independently among branches, with each branch rate sampled from independent identical log-normal distributions. The prior on the mean of the substitution rate was modelled as a gamma distribution with the rate parameter reflecting the mean distance across all genes between *Amborella trichopoda* and *Vitis vinifera*. Assuming a divergence time equal to 210 Ma (Barba-Montoya *et al.* 2018; Morris *et al.* 2018) I estimated a rate of 0.15^{-10} substitutions per site per year. Following Morris *et al.* (2018) I fixed the shape parameter at 2 (a diffuse shape) which provides a scale parameter of 13.

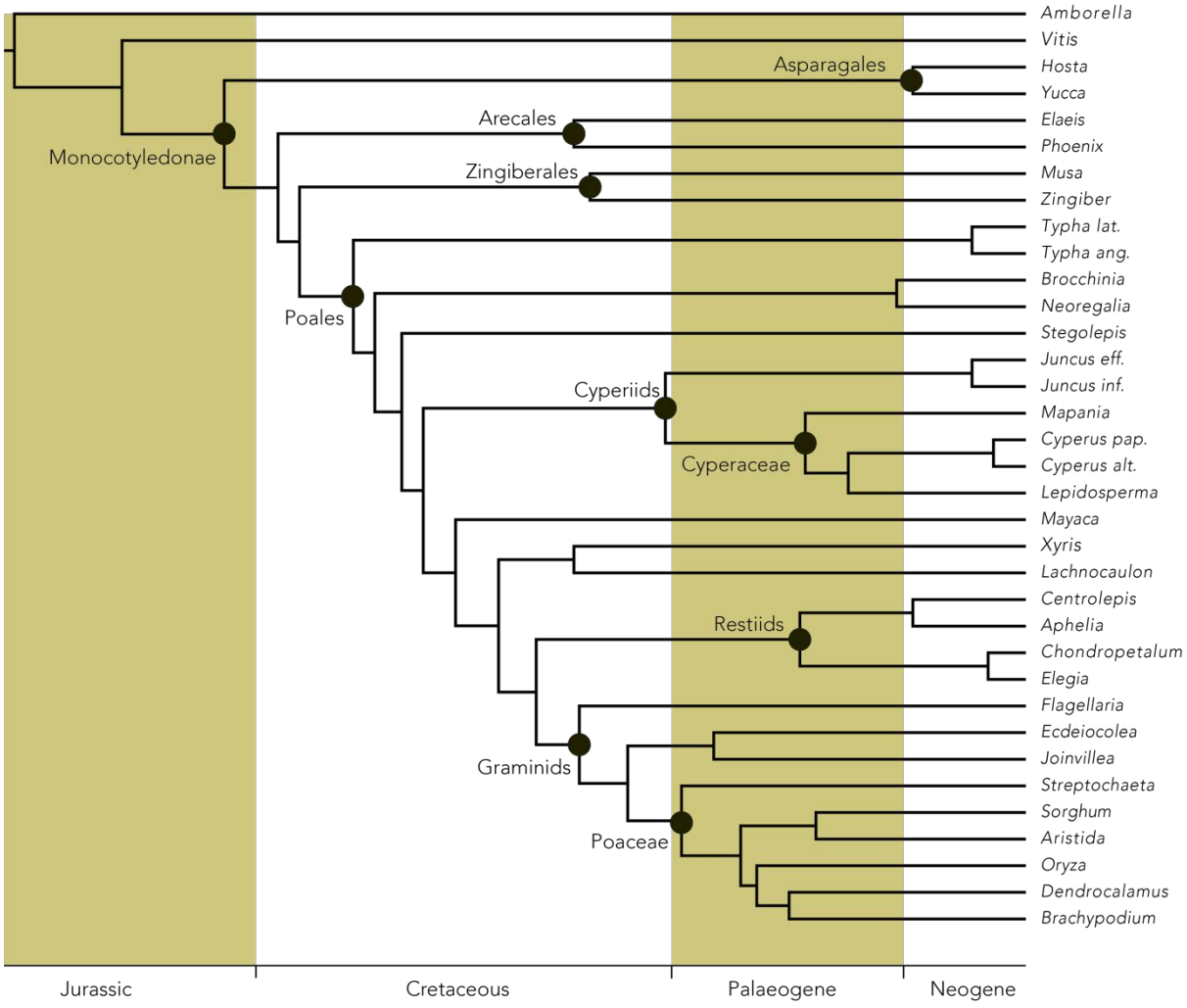


Figure 5.1. Phylogenetic relationships and divergence times of species used for the gene tree reconciliation. Species relationships were constrained to the topology inferred by Mckain *et al.* (2016) and divergence times were estimated from a molecular clock analysis of single copy orthologs.

3.2 Gene tree reconciliation

I used the orthogroups assembled by McKain *et al.* (2018) as a basis for multiple sequence alignment using MAFFT v.7.407 (Katoh and Standley 2013) and removed unreliable portions of each alignment using automated settings in trimal v.1.4 (Capella-Gutierrez *et al.* 2009).

The phylogenetic history of each gene family was reconstructed under the best fitting model as selected by iQtree (Nguyen *et al.* 2015), with 1000 ultrafast bootstrap replicates (Hoang *et al.* 2018). The set of bootstrap replicates was then used to reconcile each gene family against the species tree using the DTL model as implemented by ALE (Amalgamated Likelihood Estimation) (Szollosi *et al.* 2012). ALE uses the bootstrap samples to integrate uncertainty in gene tree reconstruction in order to infer the processes of duplication, loss and transfer within gene families. ALE is able to estimate the number of duplications, transfers and losses per node for each gene family and the total number of inferred gene duplications per node. I implemented ALE using both the undated tree of McKain *et al.* and the dated phylogeny (Szollosi *et al.* 2013; Szöllősi *et al.* 2015). I summarised the total number of duplications and transfers per node and per terminal branch.

While hybridisation occurs in natural plant populations, its contribution to ancient genome evolution in Poales is largely unknown. I repeated the reconciliation with the transfer parameter set to zero (t = 0) to compare the effect of allowing transfer between lineages.

3.3 Rates of diversification

I selected the largest available time-calibrated phylogeny for angiosperms (Zanne *et al.* 2014) and pruned the phylogeny to contain only Monocotyledonae, a total of 7060 taxa representing 11% of estimated species diversity (Givnish *et al.* 2010). Rate shifts were estimated using the diversification model in BAMM (Rabosky 2014). Despite sampling more than 10% of known diversity, this was not proportional for all lineages, and so the proportion of sampled diversity was calculated to the family level and this information was accounted for during the analysis. I selected a prior on the expected number of rate shifts of 50 based on a previous BAMM analysis that included monocots (Landis *et al.* 2018) and allowed the MCMC to run for 10^8 generations sampling every 10000 generations. Two independent runs were performed, each with four chains. I discarded the first 50% of samples as burn-in,

allowing subsequent analyses with ESS values of more than 200 for both the log likelihood and the estimated number of shifts.

3.4 Dating whole genome duplications

The three characterised whole genome duplication events were considered in light of the reconciliation results. Gene families that had duplications at the base of the non-alismatoid monocots (*tau*), Poales (*sigma*) or Poaceae (*rho*) were selected. Gene families were selected based on the following criteria: a) a topology congruent with both a history of WGD at the relevant node and current hypotheses of species relationships b) alignments of more than 150 amino acids c) broad taxonomic sampling present in both subtrees of the duplication node (Clark and Donoghue 2017). Each gene family was annotated to reflect the history of WGD following the methods of Clark and Donoghue (2017). I assembled three datasets: one for each WGD event containing exemplary gene families with a clear signal of the relevant genome duplication event. In order to more tightly constrain key nodes, I increased the taxonomic sampling for each gene family by BLASTing each gene family against genomic and transcriptomic data from the PLAZA 4.0 monocot database and the 1KP available data (Table 5.1) (Van Bel *et al.* 2018).

Where possible, nodes were cross-calibrated (Shih and Matzke 2013; Clark and Donoghue 2017) and the same set of calibrations were applied as identified in the species tree dating (Table 5.2). Molecular clock analyses were performed in MCMCTREE using the approximate likelihood model (Yang 2007). As with the species divergence estimation, I estimated the prior on the clock rate and in each case applied a gamma distribution with a shape parameter of 2 and a scale parameter of 13. The MCMC was run until convergence was reached, which was assessed in Tracer (Rambaut *et al.* 2014). The phytolith species *Changii indicum* is currently assigned to the Oryzeae (Prasad *et al.* 2011), providing a minimum age of 66.0 Ma on the divergence between the Oryzeae (*Oryza*) from the Bambusoideae and Pooideae (Iles *et al.* 2015). However, this circumscription has led to incongruence with other hypotheses about the evolution of grasses (Christin *et al.* 2014),

causing others to question this relationship and whether *C. indicum* may belong elsewhere in the BOP (Bambusoideae, Oryzeae, Pooideae) clade. To reflect this uncertainty, I repeated all dating analyses with *C. indicum* constraining first the divergence of the Oryzeae clade within the BOP clade and second the divergence of the BOP from the PACMAD clade.

Table 5.2. Maximum and minimum node calibrations used in molecular clock analyses.

Node	Minimum Age (Myrs)	Soft Maximum Age (Myrs)	Source for justification
Angiospermae – *Amborella* - *Oryza*	125.9	247.2	Morris *et al.* (2018)
Nympheales – *Nuphar* - *Oryza*	125.9	247.2	Morris *et al.* (2018)
Austrobaileyales – *Kadsura* - *Oryza*	125.9	247.2	Morris *et al.* (2018)
Canellales – *Canella* - *Sarcandra*	125.9	-	Barba-Montoya *et al.* (2018)
Eudicotyledonae – *Vitis* - *Oryza*	119.6	128.63	Morris *et al.* (2018)
Monocotyledonae – *Acorus* -*Oryza*	112.6	-	Barba-Montoya *et al.* (2018)
Alismatales – *Spirodela* - *Zostera*	96.24	-	Barba-Montoya *et al.* (2018)
Lemnoideae – *Spirodela* - *Pistia*	66.0	-	Iles *et al.* (2015)
Pandanales – *Dioscorea* - *Ludovia*	85.8	-	Barba-Montoya *et al.* (2018)
Cyclanthaceae – *Ludovia* - *Talbotia*	47.0	-	Iles *et al.* (2015)
Riponogaceae – *Colcichum* - *Smilax*	51.0	-	Iles *et al.* (2015)
Yucca – *Yucca* - *Hosta*	14.5	-	Iles *et al.* (2015)
Arecales – *Elaeis* - *Phoenix*	83.41	-	Barba-Montoya *et al.* (2018)
Coryoteae – *Phoenix* - *Sabal*	47.8	-	Iles *et al.* (2015)
Musaceae – *Musa* - *Zingiber*	74.6	-	Barba-Montoya *et al.* (2018)
Cyperaceae – *Mapania* - *Mayaca*	47.0	-	Iles *et al.* (2015)
Poaceae– *Streptochaeta* - *Dendrocalamus*	66.0	-	Iles *et al.* (2015)
Restiids – *Centrolepis* - *Chondropetalum*	27.7	-	Iles *et al.* (2015)
Stipeae – *Dendrocalamus* - *Brachypodium*	34.07	-	Iles *et al.* (2015)

4 Results

4.1 Gene and Species Tree Reconciliations

A total of 2967 gene families were reconstructed, each with 1000 bootstrap replicates. The replicates were summarised and reconciled against both the dated and undated species tree. Overall both reconciliations produced highly congruent results (Fig 5.2), though comparing the results of each gene family between the two reconciliations showed that the undated reconciliation inferred more losses while the dated reconciliation inferred a higher number of transfers (potential introgression or horizontal transfer; Fig 5.2c). The number of duplication and transfer events was calculated for each branch and in nearly all cases, the number of inferred duplications and transfers was higher in the tips of the tree. While this may reflect several lineage-specific WGD events, it is more likely a result from a large number of recent small-scale gene duplications. The number of duplication events at each internal branch supported a wave of duplication events during the evolution of the grasses (Poaceae) with high numbers of duplications at the base of Poaceae (*rho*), the BOP+PACMAD clade and the PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonoideae) clade (Fig 5.3). I also found support for the *tau* and *sigma* duplication events, though there were a greater number of duplication events on branches without confirmed duplication events, including Agavoideae, *Typha,* Bromeliaceae, Restiids, Restionaceae and *Juncus.* The number of gene transfers per branch differed from the pattern of duplication in that the number of transfers among internal branches within Poaceae were all very low. However, the high number of duplications in Typhaceae, Restionaceae, Agavoideae, *Juncus* and *Typha* all coincide with a large number of gene transfers. When gene transfer was not permitted, I observed a different pattern of duplication, with higher numbers of duplications inferred closer to the root of the tree and fewer towards the tip.

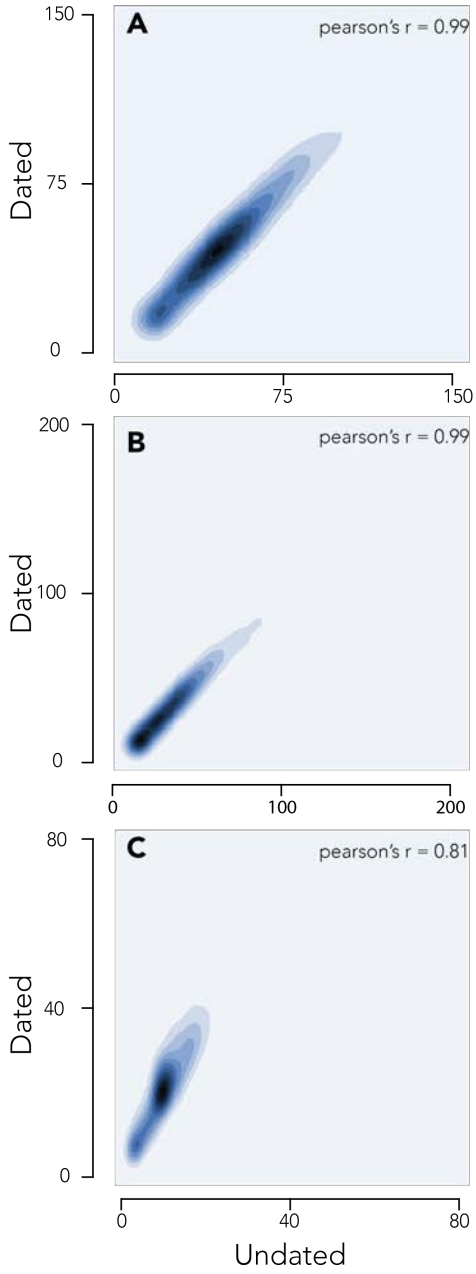


Figure 5.2. Incidences of A) Duplication B) Loss and C) transfer between gene families under the dated and undated reconciliation models. Results are highly congruent, except for higher rates of transfer in the dated model.

5.2. Rates of Diversification

The BAMM analysis identified a best fitting model which proposed 95 diversification shifts. However, the single most probable rate shift configuration could not be distinguished and so the 95% credible set of shifts, which is the set of rate shifts that explain 95% of the probability of the data, were presented (Fig 5.4). The results indicated significant rate shifts at the base of the BOP+PACMAD clade and on the branch leading to Poales (Fig 5.4). The position of each WGD event was mapped onto the phylogeny to illustrate the timing of WGD relative to any rate shifts (Fig 5.3). The highest mean rates across the tree were in the BOP+PACMAD clade and the orchids (Aparagales), although there were also notably high rates within Cyperaceae.

The mean net diversification rate (speciation minus extinction) was estimated from the posterior distribution globally for the entire dataset and for the subsequent lineages that underwent each of the three major duplications. The non-alismatoid monocots have a mean diversification rate that is indistinguishable from the background rate (Fig 5.5). However, both Poales and more emphatically Poaceae have elevated diversification rates (Fig 5.5).

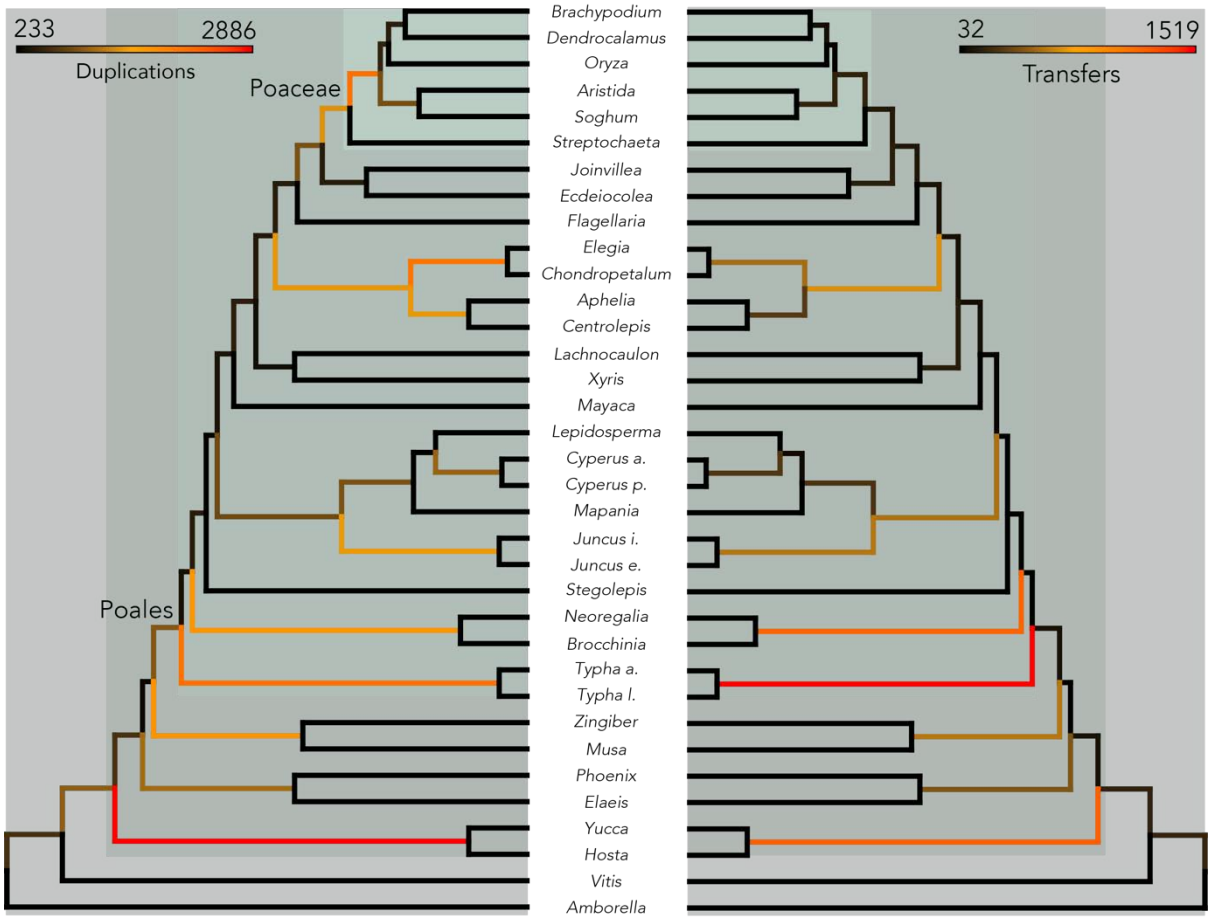


Figure 5.3. The frequency of gene duplication (left) and gene transfer (right) across all gene families inferred under the dated model in ALE. The frequencies on the terminal branches are not shown.

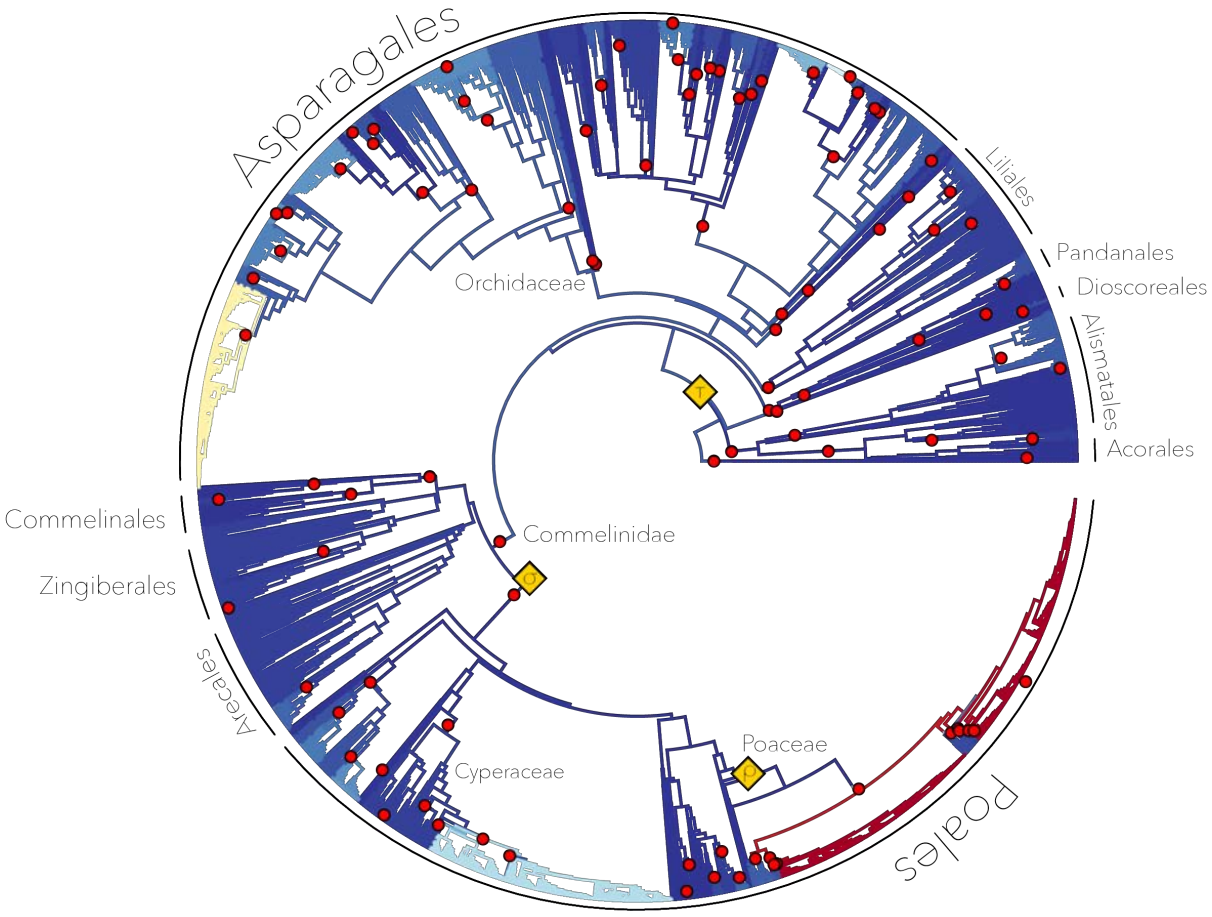


Figure 5.4. Rates of net diversification (speciation – extinction) among Monocotyledonae inferred using BAMM. Faster rates are indicated in warmer colours and slower rates in cold. The position of the putative whole genome duplication (WGD) events is displayed in yellow diamonds: *Tau* in the ancestor of non-alismatoid monocots, *sigma* in Poales and *rho* in Poaceae. Red dots indicate significant shifts in the rate of diversification from the single most probable outcome.

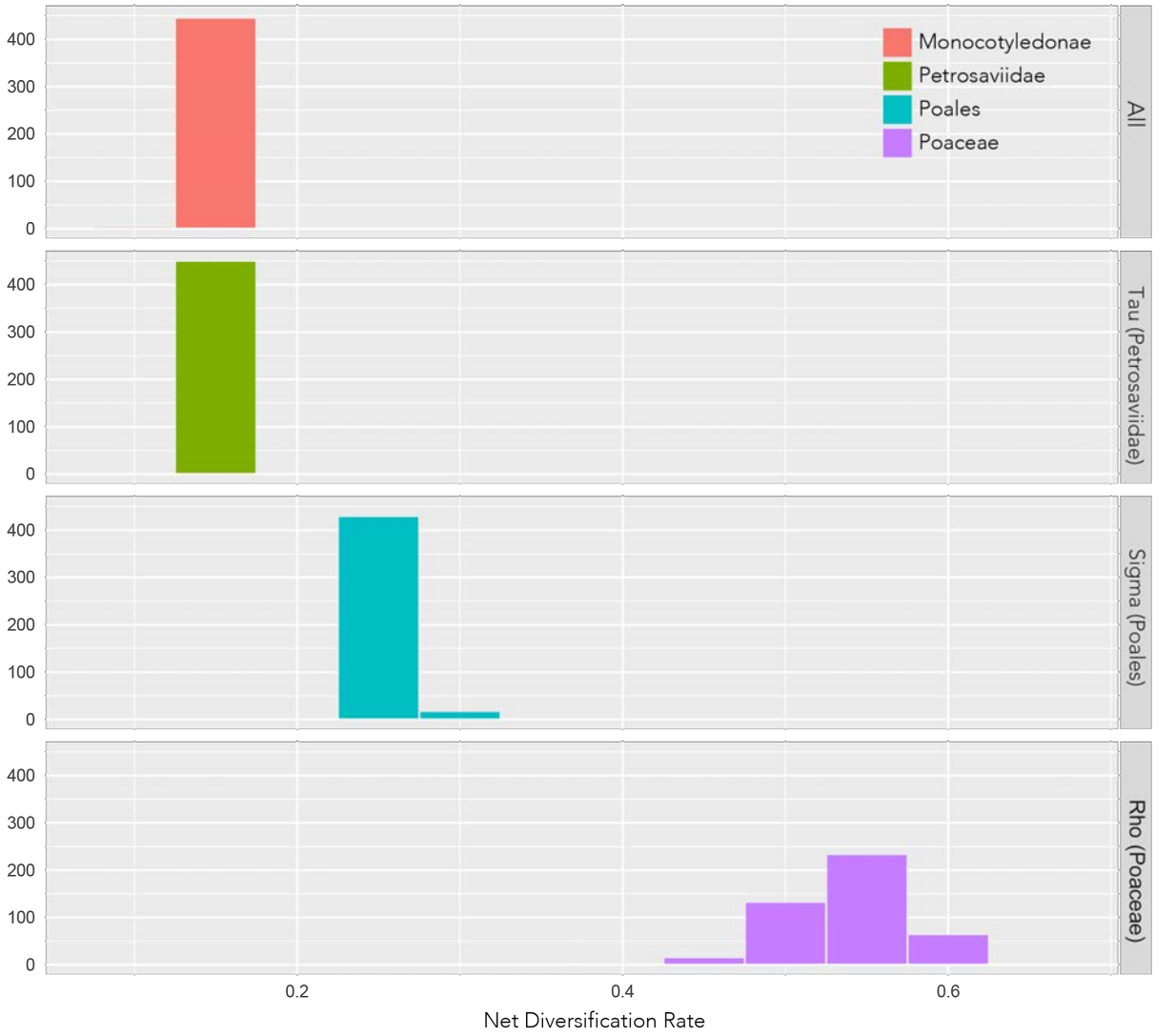


Figure 5.5. Net diversification rates across monocots and the descendants of each of the three major WGD events. Rates are averaged across 500 samples from the posterior distribution.

4.3 Inferring the age of WGD events

The criteria applied to selecting suitable gene families for molecular clock analyses resulted in datasets contained 15 (*tau*), 22 (*sigma*) and 70 (*rho*) gene families. All three analyses produced species divergence estimates that were highly congruent. The inferred ages for the *rho* event varied depending on the calibration scheme (Table 5.3), though the age estimates for *sigma* and *tau* were robust to either scheme. Under the first calibration scheme, with *C.*

indicum assigned to Oryzeae, the age of the *rho* duplication event and the origin of Poaceae is inferred to be more ancient, while assigning *C. inidicum* to crown group Poaceae allows a younger origin (Table 5.3). These results suggest that, with the exception of *rho* and the divergence of crown Poaceae, the timing of the *sigma* and *tau* events, as well as the divergence of the major lineages within the monocots is robust to the phylogenetic assignment of *C. indicum.*

Table 5.3. The inferred ages of duplication events and the immediate species divergence events that followed. All ages are provided as 95% HPD estimates.

Node	***Tau***	**Petrosaviidae**	***Sigma***	**Poales**	***Rho***	**Poaceae**
Age range (Ma) *C. indicum* derived	153-140	148-132	133–122	131-118	91–87	84-80
Age range (Ma) *C. indicum* basal	152-139	146-133	136-126	133-121	80-77	74-71

5 Discussion

5.1 Phylogenomic Methods for Inferring WGD

The use of sequence data to infer the phylogenetic timing of WGD events is dependent both on the reliability of the sequence data and the model used to reconcile the gene trees against the species tree. I demonstrated a novel application of ALE, a probabilistic method of gene-tree reconciliation (Szollosi *et al.* 2012), and have demonstrated that it is capable of recovering comparable results to previous methods (McKain *et al.* 2016). I found support for the three previously characterised WGD events *rho*, *sigma* and *tau*. As with previous approaches, I found the signal to be strongest in the youngest duplication event (*rho*; Fig 5.3), likely caused by the pattern of decay in gene retention over time. In addition to the characterised events, I also found strong support for further putative WGD events (Fig 5.3). These included Agavoideae (Asparagales) and Restionaceae, which were also supported in the analyses of McKain *et al.* (2012) and McKain *et al.* (2016), as well as a large number of duplications in the branch leading to the Restiids, Bromeliaceae and Typhaceae (Fig 5.3). These results confirm a likely scenario of repeated WGD events across the backbone of the Poales phylogeny (McKain *et al.* 2016). The exact number of WGD events is currently unknown, but greater taxonomic sampling through new large-scale sequencing projects such as the 10,000 plant genome initiative, will in all probability reveal an even greater number of ancient WGD events throughout the order, comparable to other angiosperm orders, such as Caryophyllales, which have undergone at least 27 ancient WGD events (Walker *et al.* 2017; Smith *et al.* 2018).

As an increasing number of WGD events are identified within angiosperms we stand a greater chance of understanding their evolutionary consequences. Fundamental to that is to disentangle the occurrence of auto- and allopolyploid events (Garsmeur *et al.* 2014). Hybridisation events will result in a reticulate evolutionary history which gene phylogenies coerce into a dichotomous branching pattern. After reconciliation a high incidence of gene transfer events may indicate an ancient allopolyploidy event. I found that several of the

branches with high numbers of duplication events also had high numbers of transfer events, in particular Restiids, Bromeliaceae and Typhaceae. Reconstructions of the ancestral grass karyotype and subgenome dominance have suggested that the *rho* event may have been an allotetraploidy event (Schnable *et al.* 2012), yet I found that the nodes leading to Poaceae had some of the lowest counts of gene transfer (Fig 5.3). By excluding the possibility of gene transfer from our analysis, I found that the incidence of duplication became higher towards to the root. I propose that this is an artefact caused by the inability of the reconciliation to account for phylogenetic error and instead reporting frequent duplication and high levels of loss towards the root of the tree. It also questions the biological reality of many of the transfers recorded in the previous analysis, and whether or not they simply represent an outlet for phylogenetic uncertainty.

The coupling of genome duplication and hybridisation during allopolyploidy may incur more pronounced consequences than autopolyploidy. However, presently there are too few examples of known ancient auto or allopolyploidy events to infer any difference in the evolutionary outcome, though given that there is reason to expect them to differ, future macroevolutionary studies of polyploidy would ideally be able to differentiate the two.

5.2 The Timing of WGD in Grass Evolutionary History

I was able to obtain estimates for each of the three major WGD events that are congruent with previous estimates, yet considerably more precise. The most ancient duplication event, *tau,* occurred 153-139 Ma, during the Late Jurassic – Early Cretaceous. In the wake of the duplication, several monocot orders diversified in short succession, including Dioscoreales, Pandanales, Liliales and Asparagales. The event was not accompanied by any direct increase in diversification, however, a rate shift in diversification at the base of the Commelinidae occurs within a few million years of the duplication (Fig 5.4).

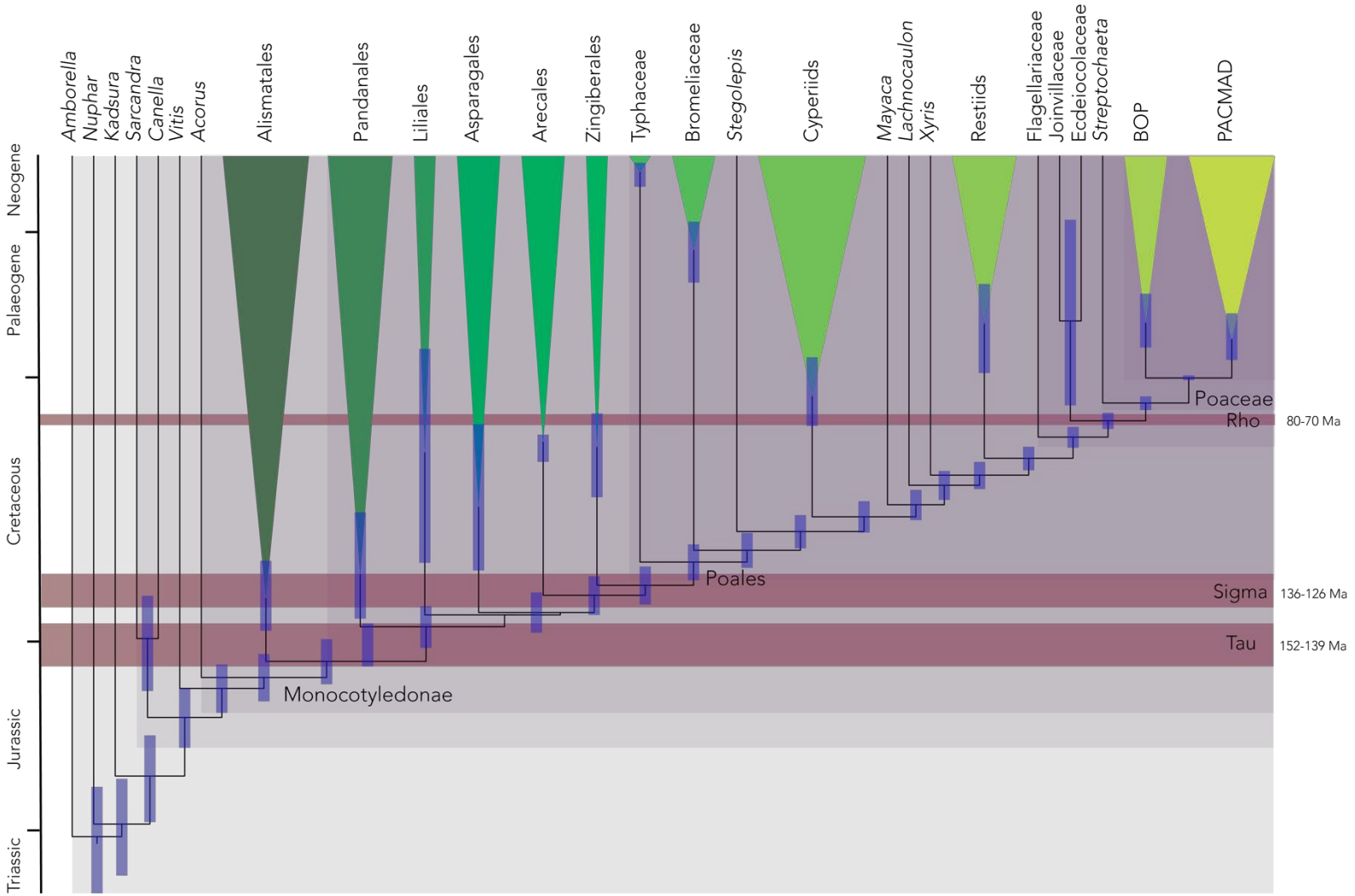


Figure 5.6. The age of the 3 major WGD event shared by the all members of Poaceae family inferred by a molecular clock analysis of multiple multi-copy gene families. Blue bars represent the 95% HPD.

The *sigma* duplication event occurred 133-121 Ma during the Early Cretaceous and preceded the divergence of the crown group Poales by 3 to 18 million years. The sigma event is coincident with an increase in diversification rate, though it should be noted that it was a very small shift and the overall rate along the branch remains low (Fig 5.4).

The timing of the *rho* duplication is more sensitive to the interpretation of the fossil phytolith *C. indicum*. Following the original assignment of the species to Oryzeae, the age of the duplication event is 91-87 Ma, during the Late Cretaceous, and the origin of the crown group Poaceae is 84-80 Ma (Table 5.3; Fig 5.6). Examining the posterior distribution of ages for the divergence of the BOP clade are very tightly constrained against the minimum, resulting in almost no uncertainty in the ages of subsequent speciation events. Following Christin *et al.* (Christin *et al.* 2014), the phytolith can also be considered as part of the broader BOP+PACMAD clade. This results in estimates considerably younger between 80-77 Ma. The age of the crown Poaceae is also younger, emerging 74-71 Ma, just before the Cretaceous-Palaeogene boundary and mass extinction event. When *C. indicum* was assigned to the Oryzeae, it produced estimates that were incongruent between the three analyses, in that age of the *rho* duplication was older than the divergence of the graminids in the other analyses (Fig 5.6). The *rho* WGD is not accompanied by any direct shift in diversification, however, the BOP+PACMAD clade shows the fastest rates of diversification within the monocots. The branch leading to the BOP+PACMAD clade showed very high rates of gene duplication, higher in fact than the *rho* branch. This suggests at least three alternative situations: a) that there was a second WGD event shared by the BOP+PACMAD clade b) that there was a burst of small-scale gene duplication on this node or that c) the position of the *rho* duplication is incorrectly placed at the base of Poaceae, and instead occurred at the base of BOP+PACMAD or that d) the reconciliation model cannot precisely locate the WGD. The first situation seems unlikely, given that syntenic analyses do not suggest two WGD events at the base of the grasses. The second hypothesis is not very parsimonious but is hard to disprove. The third situation is possible, since the syntenic analyses have not sampled any

species of the early branching grass lineages. However, our results as well as those of McKain *et al.* still support a large number of duplications at the base of Poaceae, which would be hard to explain if the WGD occurred at the BOP+PACMAD node. The fourth situation could arise through phylogenetic uncertainty, inspection of many of the gene trees showed the position *Streptochaeta* was variable. It could also be a result of Lineage Dependent Rediploidisation (Robertson *et al.* 2017), which occurs when speciation precedes complete diploidisation. If a large number of duplicates diploidised after the divergence of *Strepochaeta* but before the BOP+PACMAD divergence, then they may appear to have duplicated on that branch. The WGD event could have taken place as little as 3 million years before the divergence of *Streptochaeta* from BOP+PACMAD, which could explain this result.

5.4 Consequences of WGD in Poales

A causal link between WGD and evolutionary success has been often considered in the grass lineage. The grasses are the most dominant and successful plant family on Earth. This success has been attributed to their invasiveness ("Viking Syndrome"), and the traits that contribute to their invasiveness have been reviewed thoroughly by (Linder *et al.* 2018). Fundamental to their success is the ability to disperse and the ability to establish, which have been linked to the evolution of diaspore and the precocious embryo respectively. The diaspore among grasses is a unique and highly labile structure. Duplicates derived from the *rho* event (AP1/FUL) have been shown to be involved in the patterning of two novel bract-derived organs, the lemma and palea, which are in part responsible for the diversity of the grass diaspore (Preston and Kellogg 2007; Bouchenak-Khelladi *et al.* 2014; Linder *et al.* 2018). The grass embryo is also unique in being pre-differentiated and containing a large starch store, allowing an accelerated development (Kellogg 2000). The large starch store is linked to a synapomorphy of grasses resulting from cytosolic production of starch which is controlled by gene duplicates derived from the *rho* event (Comparot-Moss and Denyer 2009). Grasses rose to dominance by the Mid-Miocene, ~ 60 – 40 million years after the *rho* event, which implies a considerable lag between the WGD event and the eventual success of the grasses.

A possible scenario is that the *rho* event provided the evolutionary potential that later facilitated the global dominance of grasses today. Further understanding of the evolutionary development of a growing number of model systems within Poaceae will clarify the role of genome duplication in grass floral diversity (Schrager-Lavelle *et al.* 2017). A comparable developmental hypothesis relating to both the *sigma* and *tau* events is more challenging due to their more ancient nature and the paucity of model systems outside the grasses. However, early monocot evolution and the evolution of Poales are accompanied by a large number of novel traits, and so the origin of their underlying genetic controls would be of great benefit to understanding the role of WGD in morphological evolution.

7 Conclusions

The timing of WGD events in both phylogenetic and absolute terms is central to understanding their role in macroevolution. Using gene trees to infer the position of WGD events along a phylogeny requires some consideration of phylogenetic uncertainty. Here I show that ALE is a suitable method for inferring genome duplication while modelling both phylogenetic uncertainty of gene trees and a range of other genomic events. With confidence in the position of these events, I proceeded to date three successive WGD events in grass evolutionary history. Our application of the molecular clock to phylogenomic methods achieved unparalleled precision in the estimation of the age of WGD events, and further clarifies the evolutionary history of grass genomes. However, the relationship between the timing of an event and the subsequent macroevolutionary outcome remains unclear. I found no convincing support for an association between WGD and increased diversification, and the wealth of other factors affecting grass diversification are more likely drivers. While there is evidence to support the evolution of novelty as a direct result of WGD in grasses, the temporal decoupling of the event and the rise to dominance of grassland ecosystems supports a more facilitative role for WGD in grass evolution.

Chapter 6

Whole Genome Duplication as a Driver of Plant Morphological Evolution

James W. Clark, Mark N. Puttick and Philip C.J. Donoghue

1 Summary

Whole genome duplication (WGD) has been recognised as a major feature of eukaryotic evolution. The phylogenetic position of WGD events at the base of major clades, including vertebrates, teleost fishes, seed plants and flowering plants has led to speculation that WGD drives the evolution of phenotypic novelty. Undermining such speculation is the absence of formal and explicit tests that can quantify the effect of WGD events on phenotypic evolution. Such tests require knowledge of the age of a WGD, and a framework to compare phenotypic diversity among lineages that have and have not undergone WGD. I utilise morphological datasets to capture patterns of morphological evolution in the wake of WGD, using angiosperms (flowering plants) as a case study. Simulations of morphological data show that age and phylogenetic distances alone do not explain the distinctiveness and diversity of angiosperms. The WGD event coincides with high rates of morphological evolution, and traits that have arisen through WGD have made a major contribution to the morphological diversity of angiosperms. However, not all WGDs correlate with any detectable increase in disparity, and on many branches morphological evolution has accelerated in the absence of WGD. As with other macroevolutionary hypotheses that cite WGD as a driver of evolution, I find that a unifying trend is absent, and that the relationship between WGD and morphological evolution may be phenomenological.

This chapter is unpublished. The analyses were conceived by JC, PCJD and MNP. All analyses were performed by JC and the chapter was written by JC.

2 Introduction

2.1 WGD as a driver of complexity

The means through which evolutionary novelty may arise and the reasons why certain lineages appear more susceptible to innovation is one of the fundamental questions within evolutionary biology. Land plants represent one of the few lineages that have evolved towards increased complexity and a greater diversity of forms (Lang and Rensing 2015). Yet, this disparity is not explained by species diversity (Minelli 2016), or the enormous range of genome sizes within land plants (Gregory 2005).

During the 1970s, gene duplication was proposed as a suitable candidate to drive morphological innovation (Ohno 1970). By creating an extra copy of a gene, the 'free' copy would be free from purifying selection and allow it to take on a novel function. This theory was corroborated by the expansion of the major animal development regulatory genes, the *Hox* clusters, by gene duplication (Wagner *et al.* 2003). However, despite the widespread occurrence of duplication events in plants, the most frequent result of duplication is that one of the copies is lost or silenced (Lynch and Conery 2000), and the probability of a gene being retained post duplication is dependent on the nature of the duplication. Genes can duplicate individually, as part of small clusters, alongside entire chromosomes or as part of a Whole Genome Duplication (WGD). Genes with a role in maintaining the genome tend to remain as single copies and structural genes are more likely to be retained following local duplication events, but metabolic and developmental genes are preferentially retained following WGD (Freeling 2009).

These observations suit the current pluralistic model of evolution post-WGD (Conant *et al.* 2014). The initial retention of genes is controlled by relative dosage balance: genes which interact as part of regulatory networks are retained in order to preserve stoichiometric ratios (Veitia 2004). Over time, selection may relax and allow these duplicates to either gain novel functions (neofunctionalize) or share the functions of the parent gene (subfunctionalize) (Conant and Wolfe 2008; Teufel *et al.* 2016). This can occur due to selection

acting to maintain the total gene product across *both* paralogs, rather than either of the paralogs individually. In this way, drift may allow one paralog to evolve lower expression, so long as it is compensated by the other (Thompson *et al.* 2016). Once the expression of one paralog has decreased sufficiently, it becomes free to neofunctionalize. This model, named compensatory drift, implies that there should be some time delay between the moment of duplication and subsequent evolutionary consequences (Thompson *et al.* 2016). This delay has been characterized in terms of species diversification (Schranz *et al.* 2012), but can also apply to morphological evolution.

Transcription factors may play a key role in post-WGD expansions in disparity. The increase in land plant complexity is associated with an expanded repertoire of transcriptionally active proteins (TAPs), including transcription factor (TFs) and transcriptional regulators (TRs) (Lang *et al.* 2010). An increase in the number of TFs is associated with an increase in the number of cell types and by proxy morphological complexity (Wilhelmsson *et al.* 2017). Land plants show an increase in the number of TFs throughout their evolution, while numbers of TRs remain relatively constant (Wilhelmsson *et al.* 2017). TFs are preferentially retained post-WGD and their expansion has been driven by multiple rounds of WGD throughout plant evolution. The functional diversification of TFs provides a likely mechanism to explain how WGD may give rise to increased morphological disparity. For example, certain classes of TFs, including those that have a role in determining the identity of floral organs, have already been implicated as a means for WGD to have shaped the evolution of angiosperms (Chanderbali *et al.* 2016; MacKintosh and Ferrier 2017).

2.2 WGD and Morphology: A Comparative Approach

A comparative study of morphological change post-WGD has never been undertaken in land plants, largely due to the relative difficulty of quantifying morphological evolution on macroevolutionary scales. Recent comparative analyses have revealed that there is no general rule to unite WGD and diversification (Landis *et al.* 2018). Some angiosperm lineages do

appear to undergo a shift in diversification that directly coincides with WGD, other lineages after variously defined lag periods (Tank *et al.* 2015; Landis *et al.* 2018), and others demonstrate no measurable change in diversification rate

Central to a comparative study of morphological evolution is a framework that can quantify morphological disparity between lineages. Clark *et al.* (Chapter 4) quantified the relative position and diversity of each major land plant clade. Morphological evolution in plants reflects a pattern of repeated key innovations providing 'jumps' through morphological space. The differences between the major lineages far exceed the differences within them, and the characters that typify these jumps are readily recognisable. The fundamental lineages of flowering plants showed less disparity than anticipated, yet the flowering plants occupy the most extreme area of morphospace and contributed the largest proportion of the total morphospace (partial disparity). Thus, the flowering plants are highly morphologically distinct from other plant lineages.

Analyses of plant morphospace occupation are of fundamental interest, but the real value of such a space is to test hypotheses about plant macroevolution. However, our understanding of the evolution of morphological disparity is limited and there are as yet few theories to explain the distribution of disparity between lineages. Therefore, to test whether a specific event has resulted in increased disparity, it is important to identify a 'neutral' or 'background' pattern of disparity driven primarily by the order and timing of lineage divergence.

Here I present a comparison of the rates and patterns of morphological evolution across angiosperm duplication events. Molecular evidence suggests that angiosperms are at least 210-250 million years old (Barba-Montoya *et al.* 2018), with a stem lineage extending back ~ 360 million years. The evolution of flowering plants coincided with a WGD event (*epsilon*) which is evidenced in the genome of living taxa. Molecular clock estimates derived from gene families place the event between 326 and 282 Ma, early during the evolution of the stem lineage (Clark and Donoghue 2017). Other major angiosperm lineages are also characterised by WGD events, including the grasses, the majority of monocots, and the core

eudicots (Jiao *et al.* 2012; Jiao *et al.* 2014). These events have variously been hypothesised to have resulted in increased morphological diversity in the resulting lineages yet no means for framework exists for the testing of such hypotheses. I compare the evolution of morphological disparity to the rate of evolution of families of transcription families and provide the first comparative study of morphology in the wake of whole genome duplication. I show that in the case of the angiosperms, an initial WGD event produced a measurable and significant effect on the evolution of disparity. Within the angiosperms, the relationship is less clear, and across the land plant phylogeny there both multiple instances of WGD without morphological diversification and morphological diversification in the absence of WGD. I show that TF expansion does correlate with morphological complexity, but that rates of TF evolution do not correlate with known WGD events.

3 Materials and Methods

3.1 Simulating Morphological Disparity

The morphological matrix of Clark *et al.* contained 548 characters coded for 248 extant taxa. The matrix was comprised of 75% binary and 25% multistate characters. To measure the extent of homoplasy within the matrix I calculated the retention index (RI) and consistency index (CI) against a tree based on current hypotheses of species relationships (Puttick *et al.* 2018). The CI measures the amount of homoplasy within a dataset for a given tree, with the RI also accounts for the ability of synapomorphies to explain a tree. Fossil taxa that could be phylogenetically placed (Puttick *et al.* in prep) were subsampled from the matrix. The ages of nodes were constrained *a posteriori* using divergence time estimates from the literature (Laenen *et al.* 2014; Barba-Montoya *et al.* 2018; Morris *et al.* 2018) and the ages of unconstrained nodes were estimated using the equal model as implemented in the *paleotree* package in R (Bapst 2012). The mean tree length was estimated from the posterior distribution of a Bayesian phylogenetic analysis of the morphological matrix, performed

under the Mkv+Γ model implemented in MrBayes v.3.2.6 (Ronquist and Huelsenbeck 2003), and the tree length was used to inform the rate prior on the simulation model.

100 morphological matrices were simulated along the tree under the Mkv+Γ model as implemented in the R package *dispRity* (Guillerme *et al.* 2018), following the methods of Puttick *et al.* (O'Reilly *et al.* 2016; Puttick *et al.* 2017). I set the number of binary and multistate characters to match that of the empirical dataset and included the maximum proportion of non-applicable characters (50%) with equal proportions of both the 'clade' model and the 'character' model as implemented by *dispRity*. The two models differentiate character inapplicability caused by evolutionary history (clade) and character definition (character). I simulated 548 characters with the gamma parameters set to produce Consistency Index and Retention Index values that were comparable to the original dataset.

Disparity within simulated matrices was estimated following the methods of Clark *et al.* (unpublished). The matrices were first coded such that non-applicable (NA) and missing (?) were treated as separate character states (NA = "0", missing = "?"). The distances between taxa were then calculated using Gower's index, and the dissimilarity matrix was subjected to non-metric multidimensional scaling (NMDS) using the R package *Vegan* (Dixon 2003). Indices of disparity, including the mean pairwise distance (mean disparity) and the mean centroid distance (partial disparity) were calculated using the dissimilarity matrices of all 100 simulated matrices.

3.2 Rates of Morphological Evolution Across the Plant Kingdom

The rates of morphological evolution were estimated from the morphological matrix along a tree representing current hypotheses (Puttick *et al.* 2018). Rates along each branch were calculated using the discrete character rate function in *claddis*, which combines a maximum likelihood ancestral state estimation with a parsimony-based count of the number of state changes per branch and deviation from a model of equal rates (Lloyd 2016). The impact of including fossil taxa was explored by including the phylogenetically placed fossils of Puttick *et al.* (in prep.). Fossil taxa which were placed within polytomies were trimmed from the tree,

leaving 28 fossil taxa within the tree. The ages of each fossil were estimated stochastically between the first and last appearance in *paleotree* (Bapst 2012).

3.3 The non-flowering morphospace

To infer the impact of the *epsilon* WGD event, I removed all 52 characters that related to flowering from the morphological matrix leaving 496 characters. I repeated the analysis of Clark *et al.* on both the original and the subsampled matrices, using the same dissimilarity metric (Gower's index) and method of ordination (NMDS). The mean and partial disparity were estimated from both dissimilarity matrices.

To infer the effect of the evolution of the flower on the rates of morphological evolution, I repeated the discrete character rate analysis as outlined above on the non-flowering matrix. The extent of innovation within clades was calculated using Blomberg's K (Blomberg *et al.* 2003). K is a measure of phylogenetic signal where small values (<1) represent high levels of convergence between lineages, while larger values show an efficient evolution of novel character states along the tree (Clarke *et al.* 2016). I calculated K within the angiosperms for both the flowering and non-flowering datasets. I also calculated K within the seed plants, as an estimate of the influence of the flowering characters between the angiosperms and their sister lineage. For both datasets, I performed a taxonomic jackknife, randomly removing 20% of taxa for 100 iterations.

3.4 Duplication and morphology within angiosperms

Angiosperms have undergone the greatest number of WGD events within the plant kingdom, often at key positions within the phylogeny. I examined three further duplication events within the angiosperms at key phylogenetic positions. The *gamma* triplication at the base of the core eudicots, the *tau* duplication at the base of the non-alismatoid monocots and the *rho* duplication at the base of the grasses (Poaceae) (Jiao *et al.* 2012; Jiao *et al.* 2014; McKain *et al.* 2016). Cladistic matrices were sourced from the literature for each event, in each case sampling the lineages immediately either side of the WGD (Nandi *et al.* 1998;

Soreng and Davis 1998; Petersen *et al.* 2016). I randomly resolved polymorphisms within each matrix and following the methods of Clark *et al.* coded the matrix such that non-applicable data was represented by a '0' and missing data by an 'NA'. Each matrix was subjected each matrix to the previously outlined disparity analyses, each within a phylogenetic framework based on a consensus phylogeny inferred from the literature.

3.5 Rates of Transcription Factor Evolution

The total number and numbers by TF family were taken from Wilhemsson *et al.*, based on a survey of available genomes and transcriptomes (Wilhelmsson *et al.* 2017). I first looked at the total number of TFs per species. I log-transformed the totals and estimated the Brownian variance and Pagel's lambda using the motmot.2.0 package in R (Thomas and Freckleton 2012). Based on a low variance ($\sigma^2 = 0.003$) and strong phylogenetic signal ($\lambda = 0.91$), I modelled the rates of TF evolution as a continuous character using the Variable Rates model in BayesTraits v3 (Pagel 1999) with 20,000,000 generations discarded as burnin and 2,000,000 generations saved for analysis. I used the median branch length scalar as representative as the rate of evolution per branch and took the proportion of the chain during which a shift was estimated to indicate support for a shift on that branch. I repeated the above steps for each individual TF family where $\lambda > 0.75$ and $\sigma^2 < 0.01$, however I pruned the tree to remove non-Embryophyte taxa, which were not represented among many of the TAP families and calculated the mean value for three groups which in preliminary analyses showed extreme values (*Camelina* + *Capsella*, Euphorbiaceae, Triticaceae).

4 Results

4.1 Simulating Morphological Disparity

The distribution of CI and RI values for the simulated matrix were comparable to the empirical matrix. When an equivalent CI was estimated, the RI value was consistently too low, and I was not able to simulate a matrix which possessed simultaneously such a low CI

value and high RI value. As a result, I presented two classes of simulated matrix, one with a comparable RI value, and one with a comparable CI value. The total dissimilarity between the two datasets is difficult to compare, as the simulated datasets contained less variance along each axis than the empirical dataset.

As with the empirical morphospace, the shape of the produced ordinations maintained the distinctiveness of the major clades and both the simulated and the empirical morphospaces are contained between the limits of the most derived (angiosperms) and earliest-branching (Chlorophyta) lineages (Fig 6.1). In both ordinations, the charophycean algae occupy the largest total area of morphospace. The empirical morphospace shows greater distances between lineages, and a greater contribution from the seed plants and angiosperms while the simulated morphospace was dominated by older lineages. The relative contribution of each clade to the total occupation of morphospace differed between the simulated and empirical morphospaces (Fig 6.2). Some lineages show equivalent levels of disparity between the analyses, such as the bryophytes, while other lineages are less disparate in the empirical dataset, such as the lycophytes.

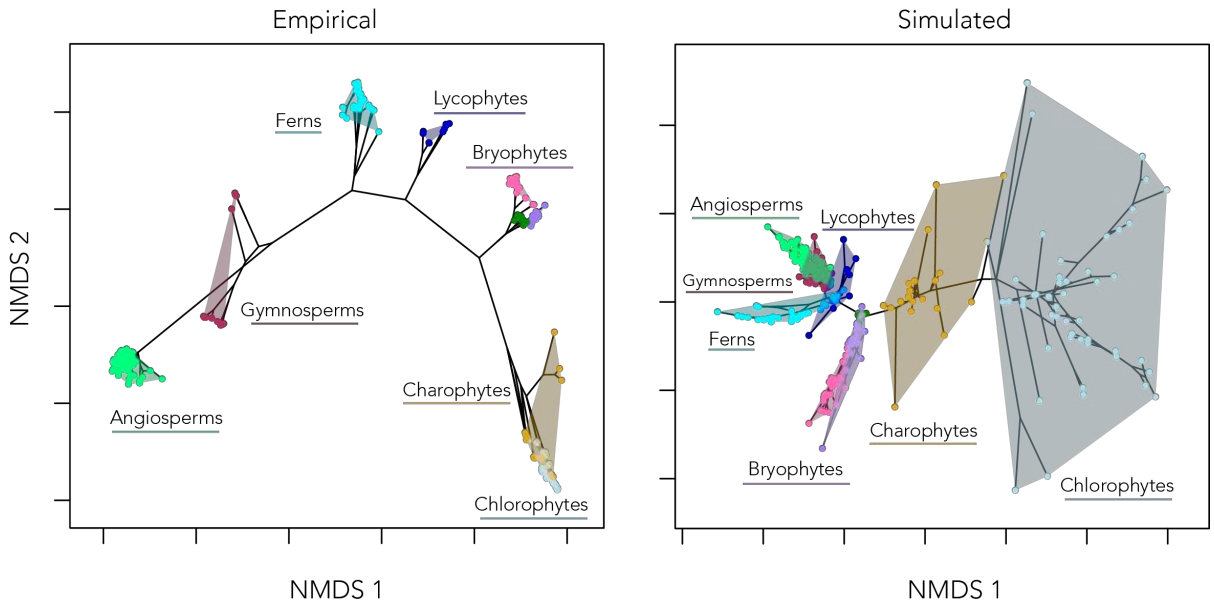


Figure 6.1. Simulated and empirical morphospaces. The empirical morphospace (A) was constructed using the dataset of Clark et al. (Chapter 4), which covers all major lineages of land plants. A consensus tree was built and dated from the literature and used to simulate (B) a morphological matrix with the same dimensions as the empirical morphospace. Both morphospaces were built based on a distance matrix calculated using Gower's dissimilarity index and non-metric multidimensional scaling (NMDS) constrained to two axes. The colours are used to indicate the major lineages of land plants.

4.2 Rates of Morphological Evolution

A comparison of the rates of morphological evolution across the tree revealed many branches deviating from an equal rates model. Compared to the null equal-rate expectation, a total of 85 branches showed a significantly higher rate of morphological evolution, and 115 branches showed a lower one. Most branches with a significant high rate were found within the angiosperms, and most low rate branches were found within the bryophytes (Fig 6.3). The highest rates were along the euphyllophyte stem branch. Significantly high rates were also observed on the embryophyte, tracheophyte, spermatophyte and angiosperm branches, as well as the branches leading to the gymnosperms, monocots and eudicots, leptosporangiate and eusporangiate ferns.

The inclusion of fossils within the morphospace resulted in different rates at several branches. It also effectively shortened the branch leading to the angiosperms (Fig 6.4), resulting in a higher rate of morphological evolution (1.72 character state changes myr^{-1}) than seen with the extant alone (Fig 6.4).

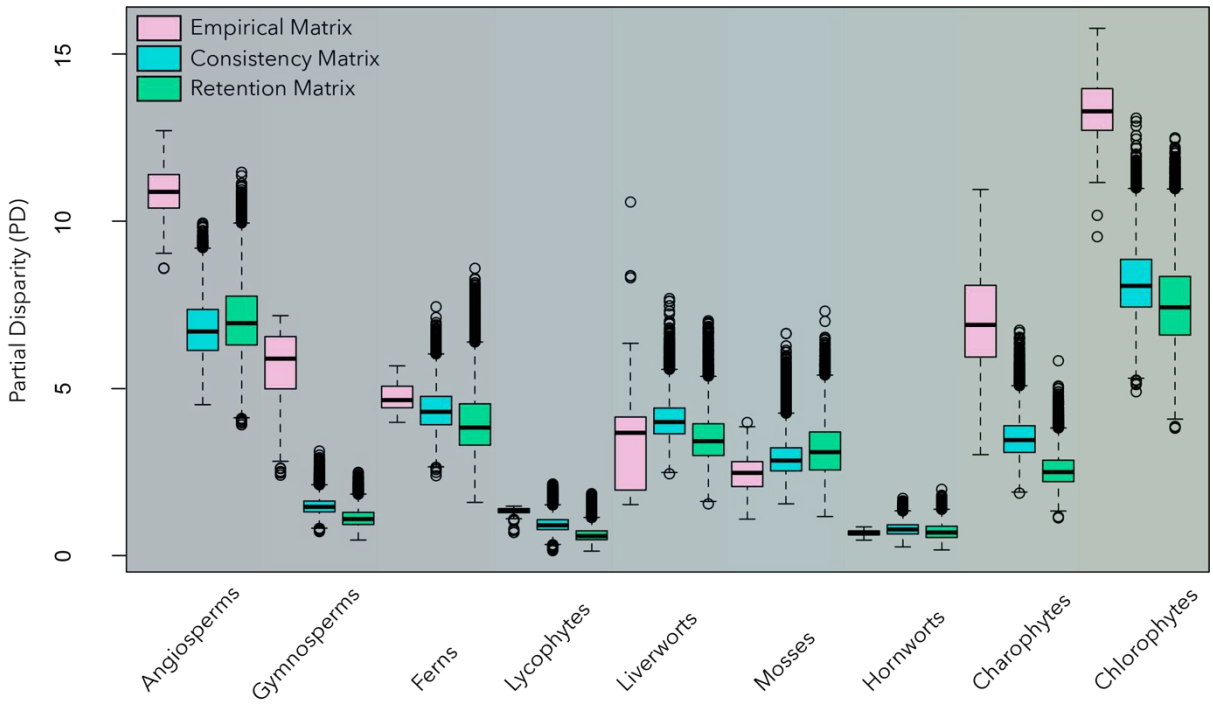


Figure 6.2. The partial (PD) and mean (d) disparity of each of the major lineages within the empirical morphospace (left, pink), consistency index-matched matrices (middle, blue) and retention index-matched matrices (right, green). The disparity indices are calculated from the distance matrix, which was calculated using Gower's dissimilarity index.

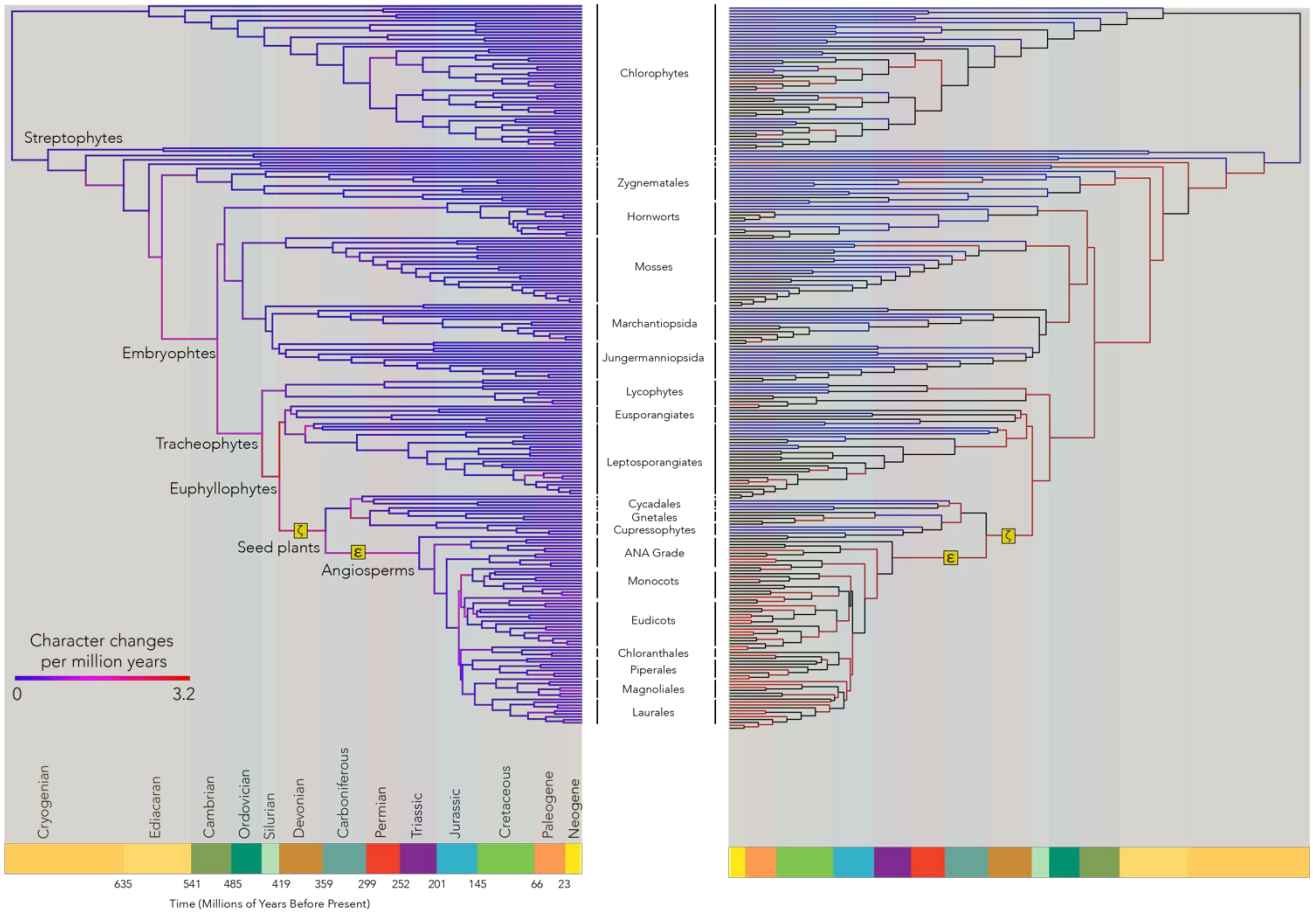


Figure 6.3. Rates of discrete character evolution across the land plants. The rates were estimated from 547 characters along a consensus tree where the topology and node ages were based on current hypotheses. On the left the deviation from a model of equal rates results in significantly high (red) and significantly low (blue) rates. The mean rate per branch is shown on the right, with higher rates in red and lower rates in blue. The phylogenetic and geological timing of the seed plant (*sigma*) and angiosperm (*epsilon*) WGD events are shown in yellow boxes.

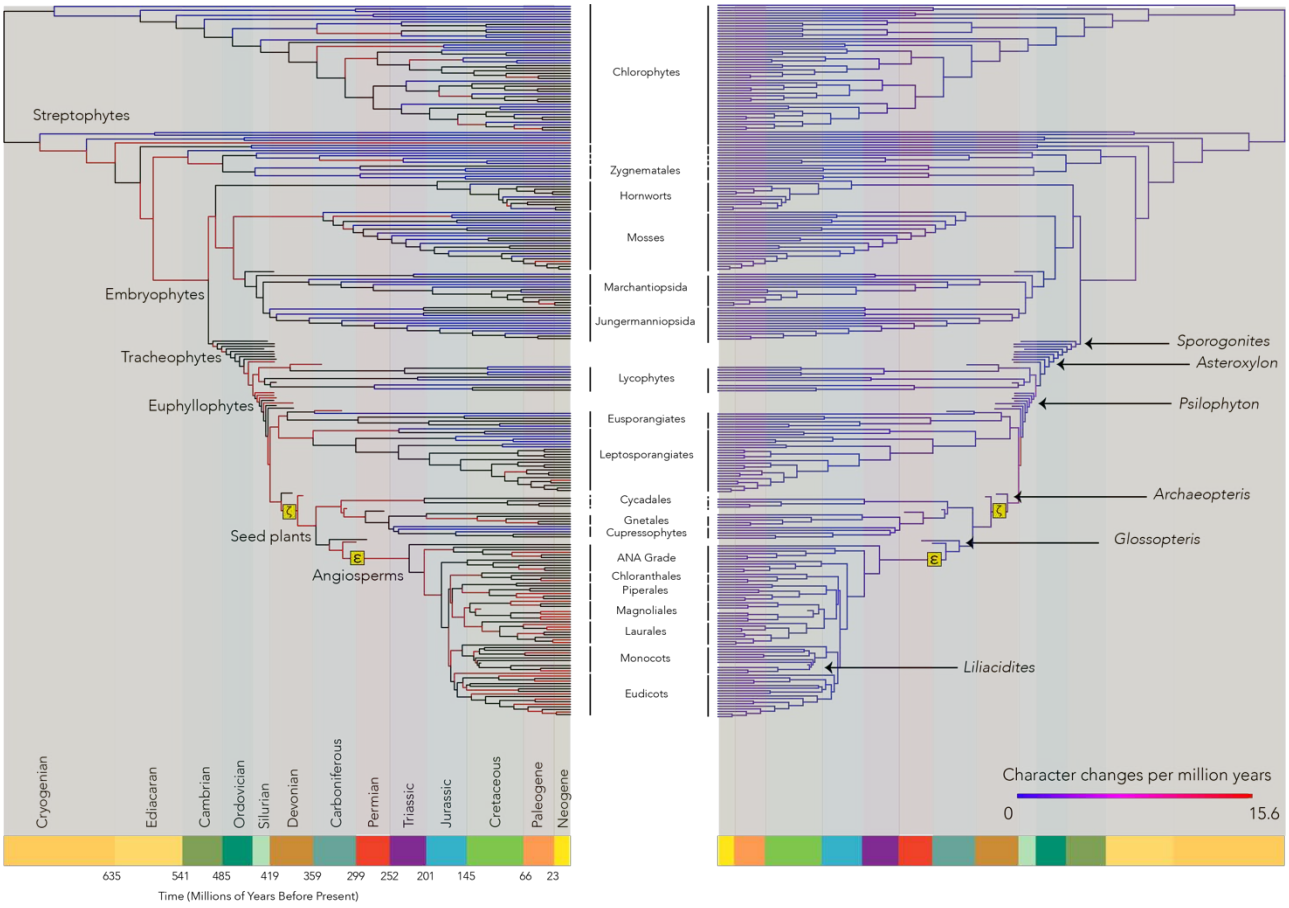


Figure 6.4. Rates of discrete character evolution across land plants including phylogenetically placed fossil taxa. On the left the deviation from a model of equal rates results in significantly high (red) and significantly low (blue) rates. The mean rate per branch is shown on the right, with higher rates in red and lower rates in blue. The phylogenetic and geological timing of the seed plant (*zeta*) and angiosperm (*epsilon*) WGD events are shown in yellow boxes.

4.3 The non-flowering morphospace

The morphospace created without the floral characters produced an overall similar pattern, yet there were both qualitative and quantitative differences in the position and disparity of the angiosperms (Fig 6.5). The distances between the major lineages of land plants were preserved, though the pronounced dissimilarity between the flowering and non-flowering plants appeared to decrease. Inferring dissimilarity from ordination spaces can be difficult,

especially in a non-metric space, yet the mean disparity and the partial disparity were both lower in the non-floral distance matrix (Fig 6.6).

Rates of character evolution in the non-flowering matrix showed that the branch on which epsilon occurs still has a significantly elevated rate of morphological evolution (0.67 character state changes myr^{-1}), however the rate is lower than with the floral characters included (1.07 character state changes myr^{-1}).

Innovation within the angiosperms did not differ between the two datasets (Fig 6.6c). Within the seed plants the distribution of K values also overlapped, although the non-flowering dataset produced a much wider range of values, indicating that innovation without flowers is much more sensitive to taxonomic sampling.

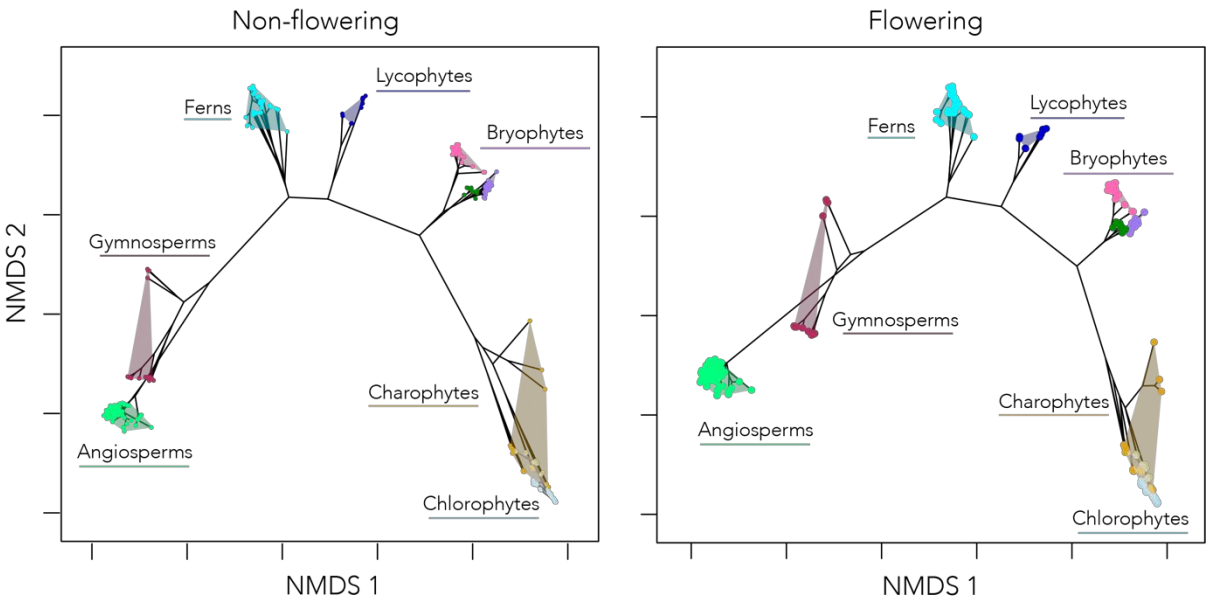


Figure 6.5. The non-flowering morphospace. The character set of Clark *et al.* (Chapter 4) was subsampled to remove 49 flowering characters (non-flowering) as a means of quantifying the impact of the *epsilon* duplication. The resulting distance matrix, calculated based on Gower's dissimilarity metric, was subjected to NMDS.

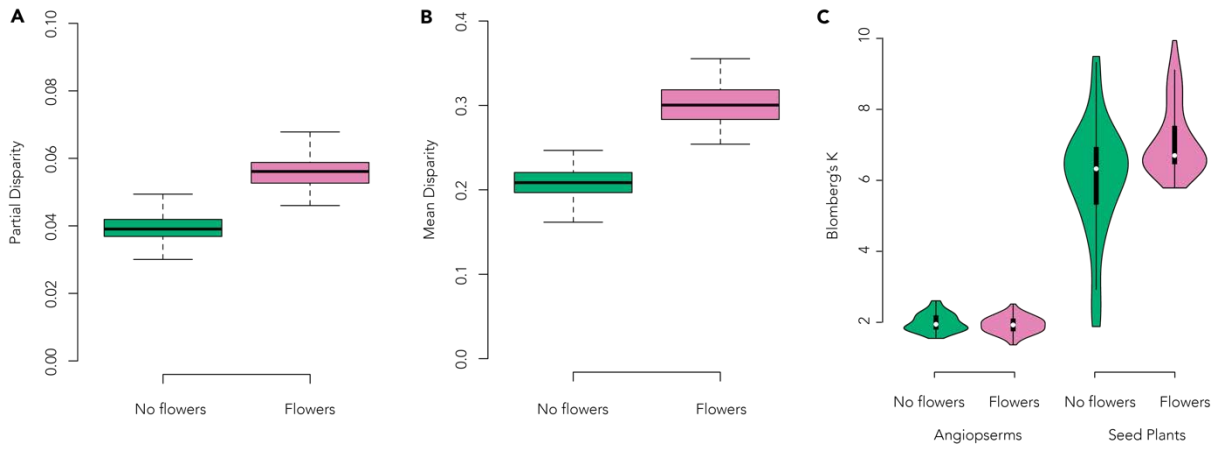


Figure 6.6. A comparison to the flowering morphospace shows lower mean (A) and partial disparity (B) in the angiosperms when the flowering characters are not considered and (C) the amount of innovation measured using Blomberg's K between the non-flowering and flowering datasets.

4.4 Duplication and morphology within angiosperms

Three independent morphospaces were ordinated comparing the morphological diversity either side of three duplication events. Poaceae occupy a highly distinct region of morphospace compared to the other graminid lineages which do not share the *rho* event (Fig 6.7a). The sampling of the other graminid lineages is restricted to a single taxon per family, and so it is not possible to compare the mean or partial disparity, yet it is still possible to estimate rates of morphological evolution along branches. The non-alismatoid monocots that underwent the *tau* duplication occupy a smaller region of morphospace than the Alismatales and occupy a distinct but highly similar region of morphospace (Fig 6.7b).

The rates of character evolution within each of these datasets suggest contradicting patterns of morphological evolution in the wake of WGD. There is no significant shift in rate following the *rho* (Fig 6.8) or *tau* (Fig 6.9) duplication event, though there is a high rate leading to the BOP+PACMAD clade.

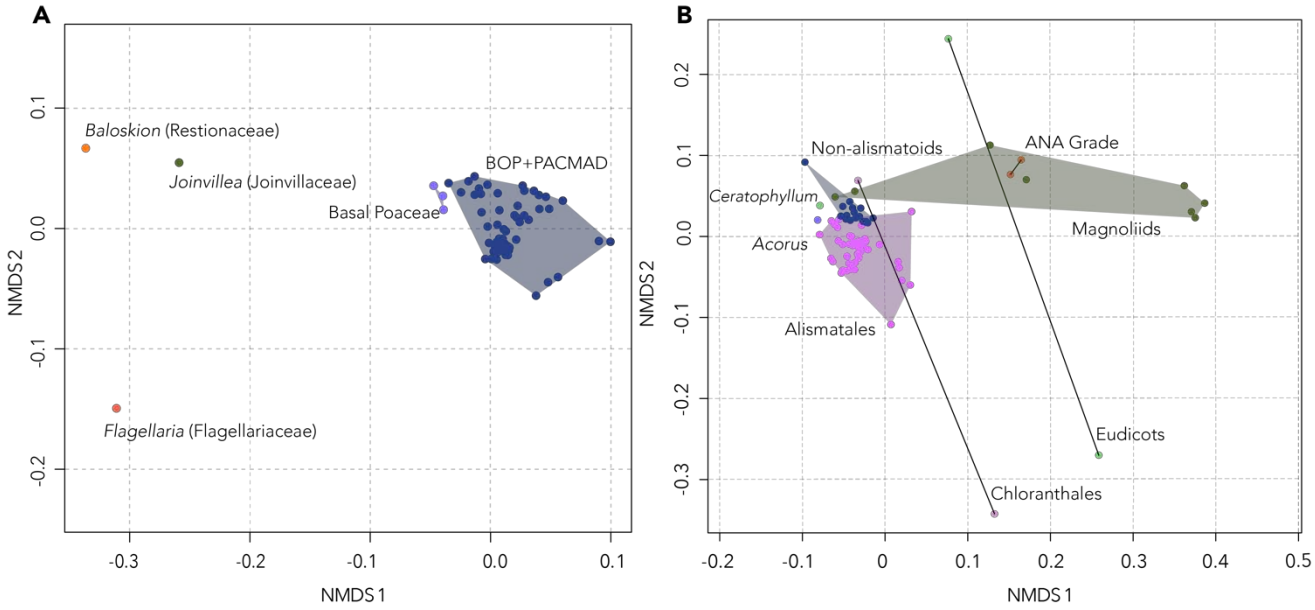


Figure 6.7. A) Morphological evolution in the wake of the *rho* WGD event, shared by all grasses (Poaceae) and B) Morphological evolution in the wake of the *tau* WGD event, shared by all non-alismatoid monocots.

4.5 Rates of Evolution of Plant Transcription Factors

The rate of evolution of the total complement of TAPs per genome revealed that significant rate variation did exist across the Viridiplantae phylogeny (Fig 6.10). The fastest rates occur on the branch leading to the zygnematalean alga *Spirogyra*, nettlespurges (*Jatropha*) and the brassica *Camelina*. Among the deeper branches of the tree, I detected no significant deviation from an equal rates model.

The analyses repeated on the individual TAP families produced similar results, with little to no significant deviations occurring along the deep branches of the tree. Rate variation most frequently occurred among branches leading to lineages that are enriched for recent polyploidy, such as *Triticum* and *Brassica.* The scoring of characters within the morphological matrix includes characters describing the presence/absence of key tissues and organs as well as further characters that describe the details of these organs. The absence of an organ or tissue type results in the scoring of 'Not Applicable' for any subsequent characters. In this way, the proportion of applicable characters within the matrix is a result of the total number of organ and tissue types present in each taxon, and a suitable proxy for

complexity. The proportion of missing data was significantly related to both the number of rounds of WGD and the total number of TAPs (Fig 6.11). In addition, there was a significant relationship between the total number of TAPs and rounds of WGD, indicating that WGD is a driver of TAP expansion.

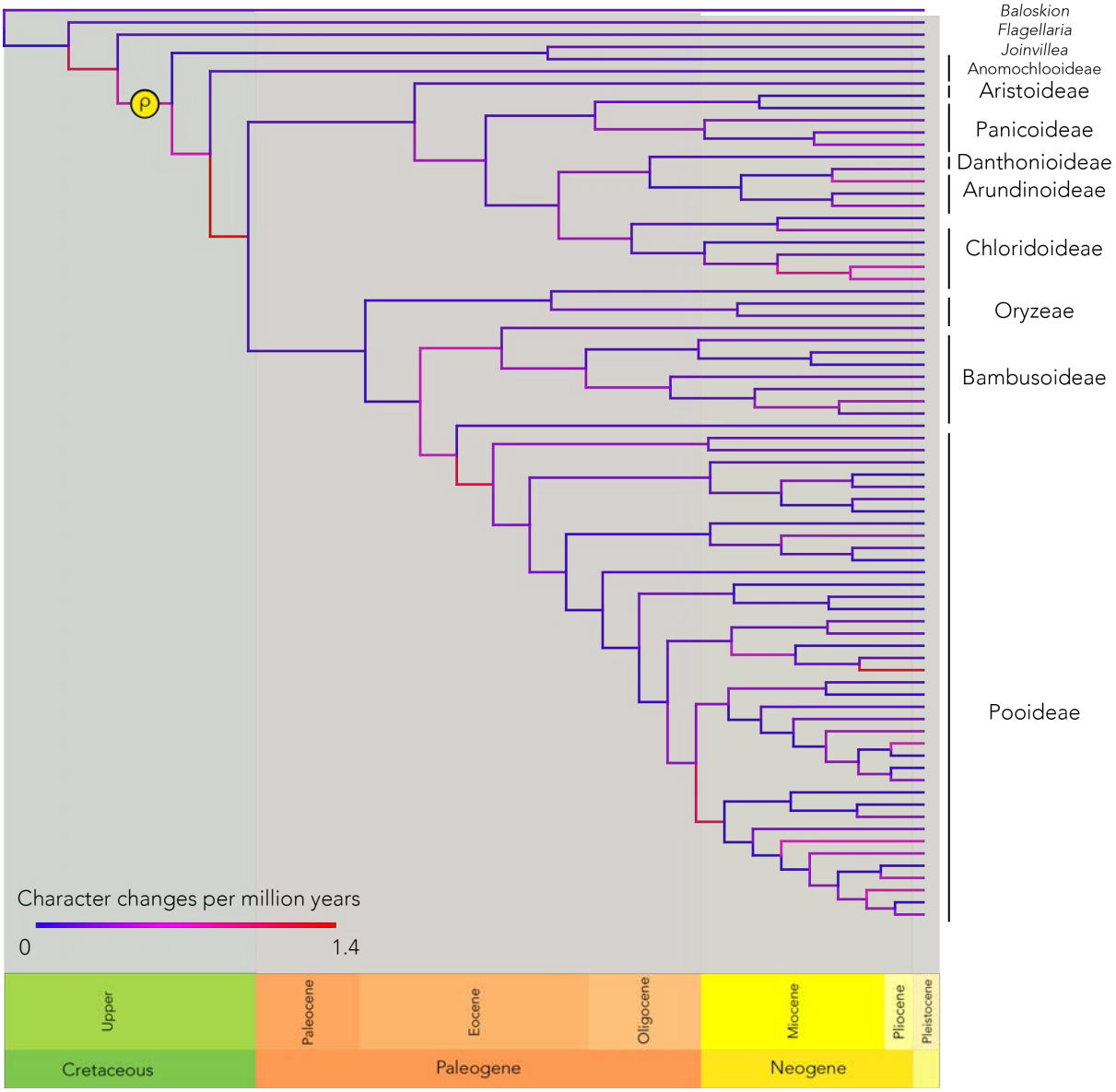


Figure 6.8. Rates of discrete character evolution across Poaceae. The mean rate per branch is shown with higher rates in red and lower rates in blue. The phylogenetic and geological timing of *rho* WGD event is shown in a yellow box.

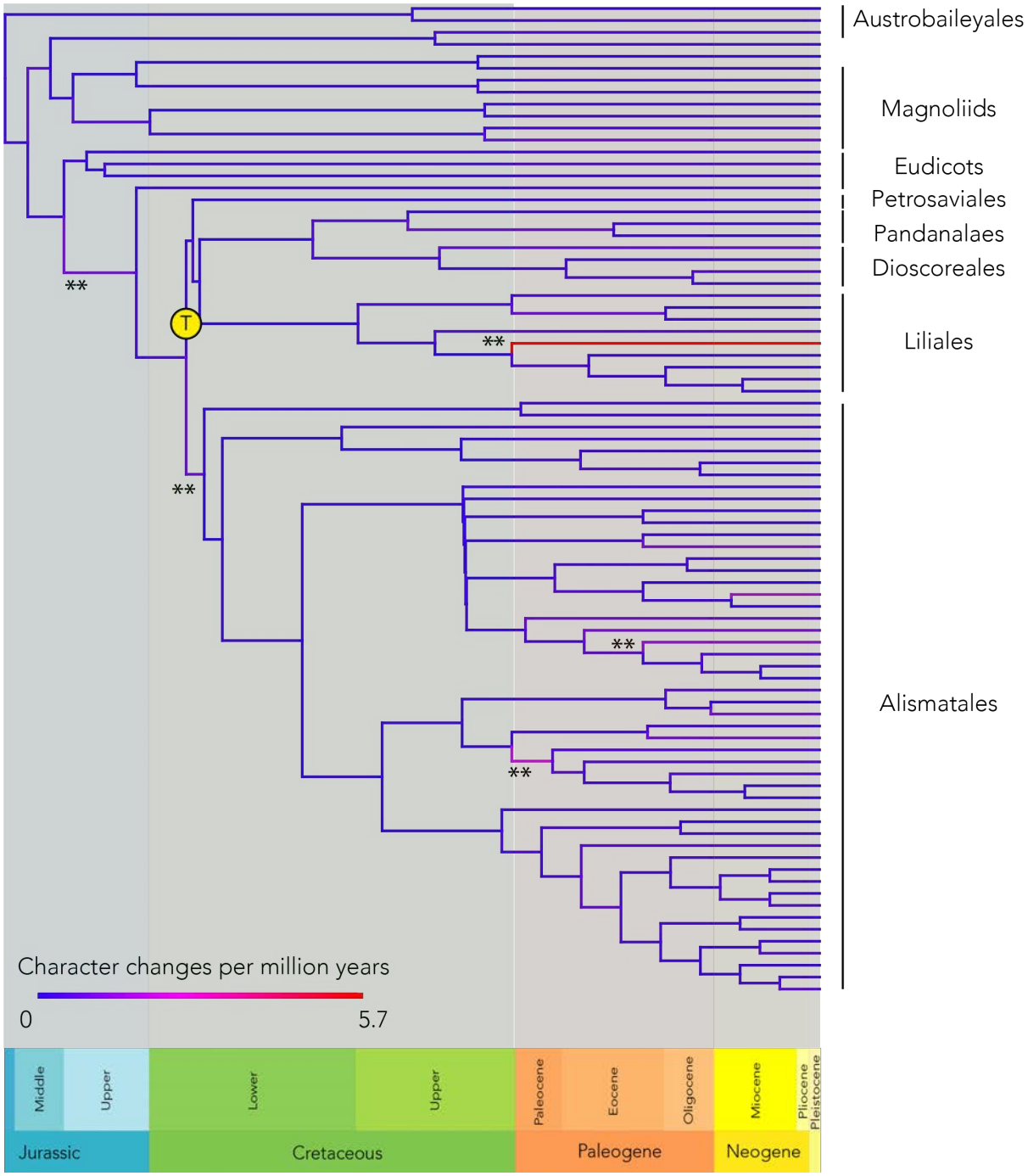


Figure 6.9. Rates of discrete character evolution across Monocots. The mean rate per branch is shown with higher rates in red and lower rates in blue. The phylogenetic and geological timing of *tau* WGD event is shown in a yellow box.

5 Discussion

5.1 Rates of morphological evolution and WGD

Simulating the land plant morphospace revealed that morphological evolution across the kingdom has not been homogeneous (Fig 6.1,6.2). These results indicate that the evolutionary history, in terms of timing and topology, alone cannot account for the contribution of angiosperms to the total morphospace. The idiosyncrasy of angiosperms within the empirical dataset is due in part to them evolving a large number of innovations; this is supported by angiosperms possessing the highest number of applicable characters. Though current methods provide a means to simulate non-applicable characters, none of the simulated matrices could precisely replicate the structure of the empirical matrix. The effect of a large proportions of contingency and matrix structure on the shape of morphospaces is an interesting question to address, since many higher taxa are defined by key innovations.

The distinctiveness of angiosperms can also be attributed to a high rate of character change (Fig 6.3). There were 1.72 character state changes per million years along the angiosperm stem branch, representing one of the fastest evolving branches across land plants. These high rates support a rapid burst of morphological evolution coinciding with the *epsilon* WGD in angiosperms. Yet other regions of the tree show similar patterns with high rates coinciding with WGD and indeed greater rates are often present on branches with no known WGD (Fig 6.3). There is an even greater rate of character evolution on the branch preceding the angiosperms, which is associated with the more ancient *zeta* duplication event ((Jiao *et al.* 2011; Clark and Donoghue 2017). Interestingly, some of these bursts of evolution appear to be an artefact of sampling only living taxa. For example, the high rate of morphological evolution leading to the tracheophytes is spread across several nodes when fossils are considered (Fig 6.4). Conversely, when fossil taxa are included it shows that some of the periods of rapid evolution were concentrated over an even shorter period of time. Most prominent among these is the branch leading to the euphyllophytes (angiosperms, gymnosperms and ferns). Newly sequenced fern genomes indicate a WGD (Li *et al.* 2018a),

yet no analyses of transcriptomic or genomic data have reported evidence for a WGD event shared among euphyllophytes. The euphyllophytes diverged more than 400 million years ago (Morris *et al.* 2018), and so it is possible that there was a WGD event but this appears unlikely with the volume of data available. Thus, the greatest burst of morphological evolution within the plant kingdom cannot be attributed to WGD.

5.2 Duplications within the angiosperms

Most lineages of land plants show evidence of ongoing lineage-specific WGD events and based on current sampling the number of ancient WGD events is highest within the angiosperms (Vanneste *et al.* 2014b). Many of these events have been shown to be decoupled from increases in diversification (Landis *et al.* 2018), but their positioning at base of major clades suggests a possible role of WGD in macroevolution. For example, the core eudicots underwent a genome triplication event (*gamma*) after the divergence of the basal eudicot lineages. This event is accompanied by an increase in diversification and the core eudicots show far greater morphological disparity than the basal eudicots (Clark and Donoghue 2018).

An additional WGD event that may have promoted morphological innovation is the *rho* event in the ancestor of all grasses (Poaceae) that took place 80-77 Ma, 3-9 million years before the divergence of the crown group. Grasses are a species rich and morphologically distinctive lineage with a unique flowering structure (Linder and Rudall 2005). The outer bracts may have formed two novel organs, the palea and lemma, and an inner whorl of bracts a further novel organ type, the lodicule. There is evidence to suggest that the MADS-box genes expressed during the patterning of these organs can trace their origins to the *rho* WGD event (Preston and Kellogg 2007). I confirm here that the grasses are highly distinct from their nearest graminid relatives and occupy a far larger area of morphospace than the remaining graminids (Fig 6.7a), supporting the view that they have morphologically diversified in the wake of the *rho* duplication.

The *tau* WGD event occurred 152-139 Ma, 19-6 Ma prior to the divergence of the non-alismatoid monocots. The event is not directly associated with any hypotheses of

morphological evolution and this is confirmed in analyses (Fig 6.7b), despite the non-alismatoid monocots being a highly diverse clade (McKain *et al.* 2016).

Empirical morphospaces are defined by the choice of taxa and characters (Wills 2001), and so a greater number of well-defined characters will lead to the distances in the morphospace approaching the true morphological distance between taxa. However, comprehensive studies of morphology are rare: most datasets tend to focus on a few traits relating to a specific functional hypothesis (for example Boyce and Knoll (2002)). Cladistic matrices can approach a more holistic view of morphology, though they suffer from the opposite problem where taxon sampling reflects a specific hypothesis. The datasets included in this study represent a move towards a quantitative test of the role of WGD in morphological evolution, yet comparisons between datasets are difficult. Without standardised characters, I cannot definitively compare the effects between different WGD events. While the *rho* WGD event does not seem to have produced a significant shift in the rate of morphological evolution (Fig 6.8), it was based on fewer characters than that used to examine *gamma*, and the taxon sampling used to study *tau* is more heavily focussed on those taxa that had not undergone the WGD event. Indeed, the nature of cladistic matrices tends to emphasise the differences within the clade of interest, which may inflate the disparity of a lineage relative to its outgroup. This could explain the low rates in non-alismatoid monocots observed here, since the original matrix was designed to estimate species relationships within Alismatales.

5.3 Floral Diversity and Morphological Disparity

A direct comparison between lineages that have and have not undergone a WGD event is problematic amongst plants, since so many lineages have undergone independent WGD events. The effects of the *epsilon* WGD event relative to the gymnosperms is difficult to disentangle, since the gymnosperms are thought to have undergone several lineage-specific WGD events (Li *et al.* 2015). An alternative approach is to examine specific outcomes of a WGD event and to quantify the resulting effect on disparity.

One of the most striking evolutionary developmental hypotheses linked to WGD in plants is the putative role for WGD in the evolution of the flower (Chanderbali *et al.* 2016). The MADS-box MIKC transcription factors control the development and patterning of the flower (Coen and Meyerowitz 1991; Theissen *et al.* 2016). The MADS-box transcription factor family extends back the ancestor of plants and animals, but several lineage specific duplication events have been linked to the evolution of plant morphogenesis (Airoldi and Davies 2012; Chanderbali *et al.* 2017). In particular, the MIKC class proteins are known to have expanded from an ancestral repertoire of 11 in seed plants to 17 in angiosperms, giving rise to the floral patterning genes (Gramzow *et al.* 2014). This expansion is attributed to the *epsilon* WGD event and has been posited as an explanation for Darwin's 'abominable mystery' (Chanderbali *et al.* 2016).

Given the causal link between the evolution of the flower and the *epsilon* genome duplication event, I examined the contribution of the flower, and by proxy the WGD event, to morphological evolution. A re-analysis of a 'non-flowering' morphospace showed that the disparity of angiosperms decreased in the absence of flowers (Fig 6.5,6.6). The distance between angiosperms and their sister lineage is qualitatively shorter, as well as a lower partial disparity. In addition, the mean disparity within the angiosperms decreased and so the *epsilon* duplication has distinguished flowers not just from their nearest relatives but from each other as well.

However, flowering plants remain distinct within the morphospace even without the floral characters (Fig 6.5). Angiosperms are also characterised by other reproductive and vegetative innovations as well as unique character combinations, including a reduction in ovule size function (Leslie and Boyce 2012), the evolution of vessels and tracheids (Boyce and Leslie 2012), leaf anatomy (Zwieniecki and Boyce 2014) and the evolution of the endosperm and fruit (Doyle 2012). This is reflected in the rate of the discrete character evolution in the non-flowering matrix, which remains significantly high. However, there is a marked decrease in the rate of character evolution, showing that much of the phenotypic evolution along the angiosperm stem was floral. These results show a complicated picture: that the

epsilon WGD event has had a considerable effect on angiosperm disparity, yet there are a large number of other innovations that characterise angiosperms which have not presently been traced back to the *epsilon* event.

Furthermore, the MIKC transcription factors are known to have neofunctionalised to regulate other aspects of angiosperm development, including the fruit (Fujisawa *et al.* 2014). As such, our estimates of the impact of the epsilon duplication event on angiosperm disparity may be conservative. Understanding the order of evolution of these traits along the stem is an outstanding question (Sauquet and Magallon 2018), and their timing relative to the duplication event, as demonstrated in in vertebrates (Donoghue and Purnell 2005), will be fundamental to refining our ideas about the role of *epsilon* in flowering plant evolution.

5.4 Dynamics of Transcription Factor Evolution

An increase in the number of transcription factors is believed to correlate with morphological and developmental complexity (Lang *et al.* 2010). Using an alternative measure of morphological complexity, I also found that, when phylogenetically corrected, both the number of TFs and the number of rounds of WGD did not correlate with morphological complexity (Fig 6.11). Transcription factors have a known role in contributing to the development and the generation of morphological disparity. The preferential retention of TFs post-WGD indicates that WGD is a means of expanding the repertoire of TFs within a lineage and so generating greater phenotypic diversity. When I modelled the rate of evolution of all TAPs across the land plant phylogeny, it agrees with previous findings that most families of TAP were present in the ancestral land plant (Catarino *et al.* 2016; Wilhelmsson *et al.* 2017). Surprisingly, there was no evidence for further rate shifts associated with the evolution of novel lineages or known WGD events (Fig 6.10).

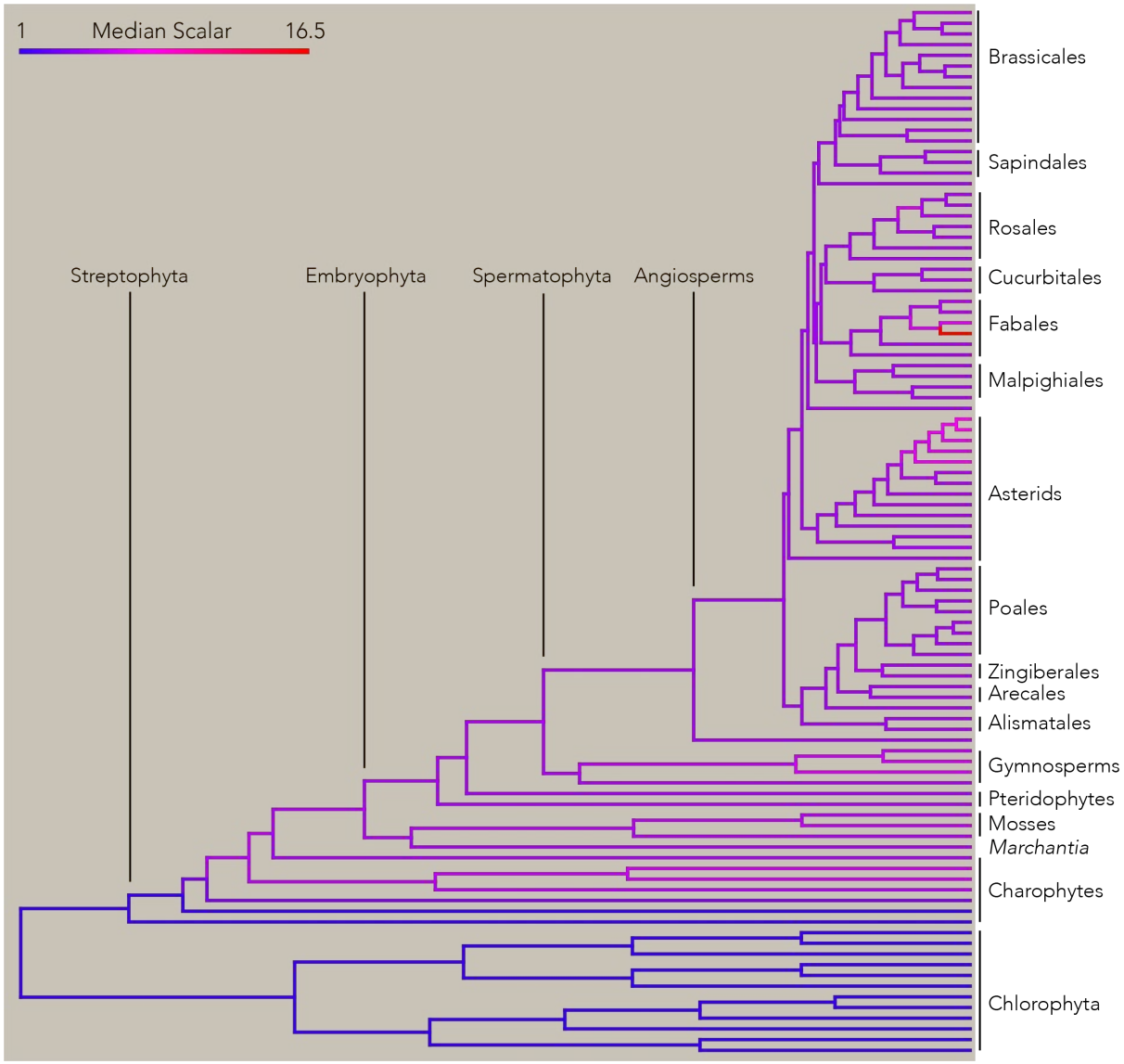


Figure 6.10. Rates of transcriptionally active protein (TAP) evolution across the plant kingdom. Rates represent the median scalar of the base rate, varying from low (blue) to high (red).

Likewise, the analyses of individual TAP families revealed that none of them have expanded in coincidence with any WGD event nor are they associated with any of the longest morphological branches. The only TF family that did show an increased rate was the YABBY family, which arose in seed plants. These findings suggest that the recurrent WGD events throughout land plant evolution have not produced any increases in the rate of TAP

family evolution. However, the major shift in the rate of TAP evolution does coincide with a high rate of evolution leading to the derived streptophytes (Fig 6.4).

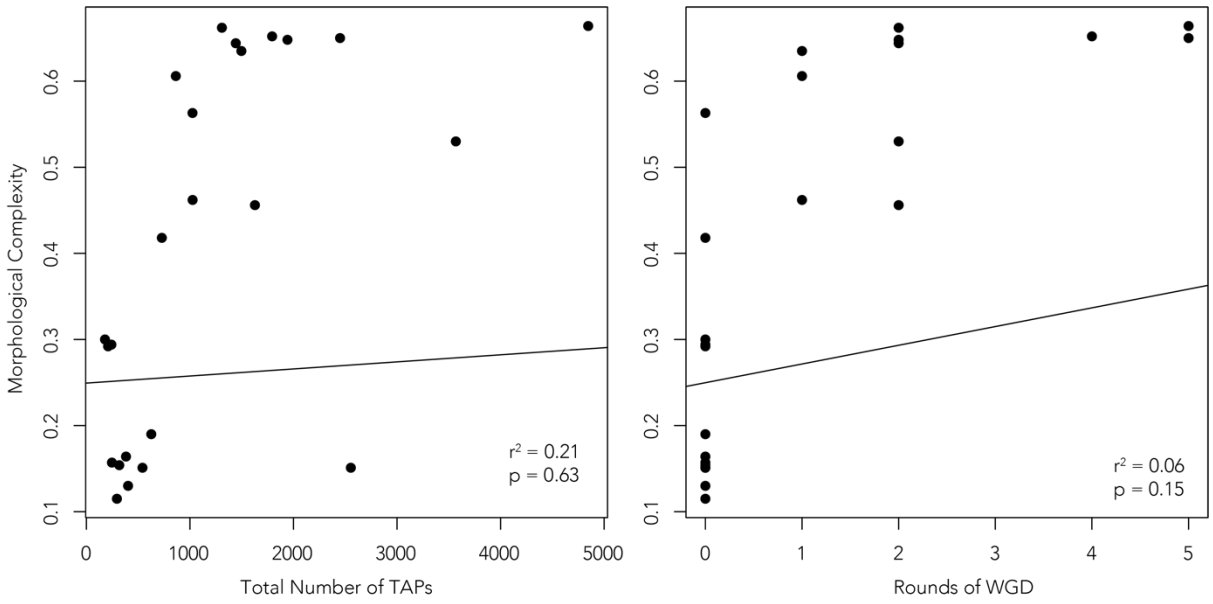


Figure 6.11. The relationship between morphological complexity (proportion of applicable characters from the matrix of Clark *et al.*) and a) total number of transcription factors and b) rounds of WGD. A phylogenetic generalised least squares (PGLS) model was fitted to the data.

6 Conclusions

The role of WGD in driving plant morphological evolution appears either lineage or context-specific. In the case on the angiosperms and the *epsilon* WGD a clear link can be drawn between WGD and an expansion into a novel region of morphospace. While WGD is not the sole cause of the morphological radiation of angiosperms, it has significantly contributed. Furthermore, these macroevolutionary observations are corroborated by evidence from evolutionary developmental studies that demonstrate that WGD-derived paralogs are fundamental for floral morphogenesis.

WGD do not always promote increases in the rate of morphological evolution. In the case of the eudicots, it seems that the contribution of the gamma triplication to floral

diversity has contributed to an increase in disparity. However, the *rho* and *tau* duplication in grasses and monocots seem to have had no measurable impact on morphological disparity. The finding that WGD seems to catalyse morphological evolution mirrors the relationship between WGD and diversification. It is possible that the relationship between genome duplication and morphological evolution is not easy to generalise. As with species diversification, it seems that some WGD events coincide tightly with a shift in morphological disparity while others appear to have no measurable effect. For example, the WGD event that occurred in the ancestor of all non-alismatoid monocots, *tau,* appears to have led to no directly measurable increase in morphology.

Further work is needed to standardise comparisons of morphology across the plant kingdom and to expand on phenotypic datasets that can test a greater number of WGD events. The evolution of transcription factors is a potential driver of morphological evolution in the plant kingdom, yet I find no evidence of increased rates of evolution associated with any known WGD events. These contrasting results currently propose more questions than they answer, namely what causes some lineages to evolve major innovations in the wake of WGD while others do not. The same question reversed is perhaps just as interesting: how do some lineages remain morphologically unchanged in the way of such large-scale genomic change?

Chapter 7

Whole genome duplication and the origin of Equisetaceae

James W. Clark, Mark N. Puttick & Philip C.J. Donoghue

1 Summary

Whole Genome Duplications (WGD) have been linked to evolutionary success across the tree of life. Thus, the discovery of a WGD in a species poor lineage represented by a sole extant genus is intriguing. However, lineage diversity is only one measure of evolutionary change and morphological distinctness may be the long-term consequence of a WGD. *Equisetum* belongs to a once morphologically and species rich lineage, the Sphenopsida. Recent fossil discoveries have painted a more informed picture of the evolutionary history of *Equisetum* and the discovery of a WGD event have renewed interest in this ancient branch of the plant tree of life. We investigated first whether the WGD event in *Equisetum* coincides with the K-Pg boundary. Second, we asked whether WGD has resulted in the evolution of increased morphological disparity and novelty within the Equisetales. We analyse all available transcriptomes from living taxa to show that the WGD event is shared by the subgenera *Equisetum* and *Hippochaete*. Contrary to previous estimates, the WGD event is incredibly ancient, coinciding with the Permian-Triassic boundary and the evolution of the stem group Equisetaceae. We show that rates of morphological evolution across the phylogeny are heterogeneous yet despite this, we find that much of the morphological distinctiveness evolved prior to the WGD and it is associated with few instances of novelty. Finally, we show that WGD correlates with an increase in genome size, all detected on one of the longest branches in the plant kingdom.

This chapter is unpublished. All analyses were conceived and undertaken by JC. Results were interpreted by JC, MNP and PCJD. The calibration of Equisetum was based on that written by Clarke et al. (2011) and the text has been updated where needed.

2 Introduction

An extant view of the genus *Equisetum* shows a morphologically distinctive yet species poor lineage. Recent discoveries have enhanced this view and show that while the 15 extant species of *Equisetum* have diverged relatively recently, their relatives extend far back into the Mesozoic. *Equisetum* belongs to an even larger and more diverse group, the Sphenopsida, which is further divided into the extinct Sphenophyllales and Equisetales. The relative paucity of species and morphological disparity observed in extant *Equisetum* masks a variety of body-plans which are observed back to the Carboniferous, and its evolutionary longevity and ecological ubiquity have marked it as a truly successful lineage (Rothwell 1996).

Recent studies and discoveries have thrown new light on the evolution of Equisetales. First, there have been a series of recent discoveries of increasingly ancient taxa with characteristics associated with crown group *Equisetum* (Stanich *et al.* 2009; Channing *et al.* 2011; Elgorriaga *et al.* 2015). These findings contradict previous analyses based solely on molecular data from extant taxa which suggested that crown *Equisetum* arose during the Palaeogene (Des Marais *et al.* 2003). Second, the variation in reproductive structures throughout geological time were explained by a modular pattern of development (Tomescu *et al.* 2017). The disparity of reproductive structures found in fossil Equisetales had been a hindrance in attempts to establish homology (Page 1972; Naugolnykh 2004). Tomescu *et al.* showed the differential expression of three modules was able to produce the observed diversity in extant and fossil taxa and thus establish homology between them. Finally, recent phylogenetic developments have changed our understanding of *Equisetum* evolution (Elgorriaga *et al.* 2018). Elgorriaga *et al.* inferred a phylogeny based on both morphological and molecular characters using a parsimony framework, and established *Equisetum* as more ancient than previously thought. Of particular significance was the finding that Equisetaceae is not nested within Calamitaceae but has instead evolved along an independent branch. They resolved the fossil taxon *Equisetum fluviatoides* as part of the subgenus *Hippochaete* and *Equisetum clarnoi* as belonging to the subgenus *Equisetum*, pushing

back the ages of both lineages respectively. They also placed several fossil taxa within the *Equisetum* crown group, confirming a Lower Jurassic origin.

These findings together have concluded that current estimates for the age of *Equisetum* based solely on molecular data from extant taxa are likely to be too young (Stanich *et al.* 2009; Elgorriaga *et al.* 2018). Molecular clock analyses have previously placed the age of the crown group in the Neogene or Palaeogene (Des Marais *et al.* 2003). However, while Elgorriaga *et al.* presented a timeline alongside their phylogeny, they did not explicitly model time in their analyses. Newly developed methods are able to accommodate morphological and molecular data whilst using age information from multiple fossil taxa (Ronquist *et al.* 2012; Heath *et al.* 2014; Donoghue and Yang 2016). Furthermore, after lengthy debate it has been shown that Bayesian methods outperform parsimony in terms of accuracy when analysing discrete morphological data (Puttick *et al.* 2017; Goloboff *et al.* 2018; O'Reilly *et al.* 2018), which means that the evolution of Equisetales has yet to be studied while considering all available evidence.

A survey of cytological features across ferns showed that the genomes of different lineages appear to have evolved along independent trajectories. Unlike other fern lineages where chromosome numbers can be extremely variable, all extant species of *Equisetum* possess $2n = 216$ chromosomes (Clark *et al.* 2016). This number is well above the average for even ferns ($2n = 121$), which typically feature high chromosome numbers (Clark *et al.* 2016), and so it was predicted that *Equisetum* were polyploids that had undergone multiple rounds of whole genome duplication (WGD).

The sequencing of a single transcriptome of *E. giganteum* showed that, like many other land plant lineages, *Equisetum* had undergone at least one round of WGD (Vanneste *et al.* 2015). By analysing the rates of synonymous substitutions (*Ks*) between duplicate pairs, the WGD was estimated to have occurred ~ 66 -100 Ma. Notably, evidence was only found for a single WGD event and it was proposed that the high chromosome numbers must have evolved through an alternative means. WGD has been presented as a means of explaining various macroevolutionary phenomena (Clark and Donoghue 2018), including diversification

and morphological innovation. As *Equisetum* are not very diverse, Vanneste *et al.* suggested that the WGD event may have contributed to the longevity of the lineage, despite estimating that the event occurred relatively recently during the evolutionary history of *Equisetum.* The clustering of several WGD events close to the K-Pg boundary (Vanneste *et al.* 2014a; Lohaus and Van de Peer 2016), including that of *E. giganteum,* has led to the hypothesis that WGD may have been a mechanism through which plants survived and succeeded in the wake of the ecological disturbances associated with the mass extinction event caused by the Chicxulub meteor impact and volcanism (Wilf and Johnson 2004; Renne *et al.* 2015).

To further understanding of the evolution of this enigmatic lineage, we present a macroevolutionary analysis of extant and fossil Equisetales in the light of an ancient genome duplication event. We use transcriptomic methods to phylogenetically place the WGD event in the ancestor of living *Equisetum* and molecular clock estimates to show that the WGD event far precedes current estimates. Combined with evidence for rate heterogeneity in morphological evolution, we show a temporal correlation between the expansion of the Equisetaceae and WGD.

Table 7.1. Taxa sampled for the total evidence and morphospace analyses. Where available, 1C-values are provided as used in the ancestral reconstruction of genome size.

Taxon	Age	C-value	Taxon	Age	C-value
Equisetum arvense	-	*14.65*	*Neocalamites sp.*	273-259	-
Equisetum bogotense	-	*20.65*	*Neocalamites arrondei*	240-210	-
Equisetum sylvaticum	-	*12.89*	*Paracalamitina striata*	283-268	-
Equisetum ramosissimum	-	*28.2*	*Cruciaetheca patagonica*	295-284	-
Equisetum scirpoides	-	*21.25*	*Cruciaetheca feruglioi*	295-284	-
Equisetum pratense	-	-	*Weissistachys kentuckiensis*	315-307	-
Equisetum hyemale	-	*26.3*	*Mazostachys pendulata*	315-307	-
Equisetum diffusum	-	-	*Pendulostachys cingulariformis*	307-299	-
Equisetum giganteum	-	*26.14*	*Palaeostachya decacnema*	305-299	-
Equisetum laevigatum	-	*25.7*	*Palaeostachya andrewsi*	315-307	-
Equisetum myriochaetum	-	*25.65*	*Calamostachys casheana*	315-307	*1.99 **
Equisetum fluviatile	-	*13.5*	*Calamostachys binneyana*	323-307	-
Equisetum palustre	-	*14.25*	*Calamostachys americana*	307-299	-
Equisetum variegatum	-	*30.35*	*Calamostachys inversibractis*	315-307	-
Equisetum telmateia	-	-	*Calamocarpon insignis*	323-299	-
Equisetum vancouverense	136-133	-	*Peltotheca furcata*	295-284	-
Equisetum fluviatoides	66-59	-	*Protocalamostachys arranensis*	350-335	-
Equisetum haukeanum	136-133	*6.85*	*Protocalamostachys farringtoni*	350-340	-
Equisetum thermale	166 -157	*6.08*	*Rotafolia songziensis*	*372-359*	-
Equisetum lyelli	145-140	-	*Hamatophyton verticillatum*	*372-359*	-
Equisetum laterale	199-164	*26.07*	*Bowmanites moorei*	315-307	*1.99 **
Equisetum clarnoi	41-38	*20.04*	*Bowmanites dawsonii*	323-315	-
Equisetum dimorphum	190-180	*10.44*	*Ophioglossum reticulatum*	-	*63*
Equisetites arenaceus	242-227	-	*Psilotum nudum*	-	*72.68*
Spaciinodum collinsii	247-242	-			

**Proxy values – a single representative was chosen of the Calamitaceae and Sphenophyllales and by proxy they were given the estimated values of* Calamocarpon *and* Sphenophyllum*, respectively*

3 Materials and Methods

3.1 Total Evidence Dating of Fossils and Extant Taxa

We used the morphological and molecular matrices of Elgorriaga *et al.* (2018) of 77 binary and multistate morphological characters and the *rbcL, atpA, atpB* and *matK* chloroplast genes (Elgorriaga *et al.* 2018). The matrix contained 49 taxa, including 17 extant and 32 fossil taxa spanning the Sphenophyllales + Equisetales as well as outgroup taxa *Hamatophyton verticillatum*, *Rotafolia songziensis*, *Ophioglossum reticulatum* (Ophioglossales) and *Psilotum nudum* (Psilotales).

A total evidence analysis was undertaken in MrBayes v.3.2.6 (Ronquist and Huelsenbeck 2003; Ronquist *et al.* 2012). A combination of tip and node calibrations were employed (Table 7.1) and a uniform distribution was placed on the root of between 451-384 million years (Morris *et al.* 2018). A relaxed clock model was implemented, with the clock rate prior set as a lognormal distribution with the mean estimated from a topological analysis to estimate the tree height (0.02 substitutions site^{-1} million years^{-1}) and a uniform birth death model was applied across the tree (Ronquist *et al.* 2012). The morphological data and each gene were partitioned separately, with molecular data analysed under the GTR+Γ model and the morphological data under the MKv+ Γ model. Four independent chains were run for 20,000,000 generations. Convergence between the chains was assessed based on the average standard deviation of split frequencies and by examining the parameters of the chain in Tracer including Effective Sample Size (target > 200) (Rambaut *et al.* 2014).

3.2 Rates of Morphological Evolution

To examine the rates of morphological evolution across the tree, we performed a morphological clock analysis using only the morphological dataset, with the tree constrained to the topology resolved by the total evidence analysis. As with the total evidence analysis, a relaxed clock model was used with the clock prior estimated from the tree height.

The rate of morphological evolution was estimated by sampling the effective branch lengths from the posterior distribution. We selected 1000 random trees from the posterior distribution and summarised the mean rate of morphological evolution per branch on the consensus tree. Branches in the consensus tree that were not represented in our subsample were not assigned a rate.

3.3 Morphological Disparity

The morphological matrix was recoded following the methods of Deline *et al.* (Deline *et al.* 2018), such that non-applicable (NA) states were coded as '0' and missing data as '?', to distinguish the two types of 'missing data' (Deline 2009). The distance between taxa was calculated using Gower's dissimilarity metric (Gower 1971),which has the desirable quality of not clustering taxa based on shared non-applicable data (Deline 2009). The distances were projected into two-dimensional space using non-metric multidimensional scaling, an ordination method that is not eigenvector-based, but that seeks to minimise the differences between the rank dissimilarity between taxa and their distance within the ordination space. We plotted a phylomorphospace using the majority-rule (50%) consensus tree from the total evidence analysis. The most likely ancestral state was reconstructed along the tree by summarising states across 1000 stochastic character maps (Huelsenbeck *et al.* 2003; Revell 2012) and was used to position the nodes within the morphospace.

Disparity metrics including the mean disparity (mean pairwise dissimilarity) and partial disparity were calculated based on the distance matrix (Chartier *et al.* 2017), rather than the ordination coordinates, using the *dispRity* package in R (Guillerme *et al.* 2018). We also used the ancestral node reconstructions to calculate the disparity through time, using the time-slicing approach and 'gradual split' model as implemented in *dispRity*.

3.4 Transcriptome Assembly

Assembled transcriptomes were collected from the 1KP dataset for *Equisetum diffusum, Equisetum hyemale, Culcita macrocarpa, Ophioglossum petiolatum, Tmesipteris parva, Selaginella*

kraussiana, Danaea nodosa and *Botrypus virginianus* and additional transcriptome for *Equisetum giganteum* was attained from Vanneste *et al.* (2015).

Paired end short reads were downloaded from the SRA archive for *Equisetum arvsense* and assembled following the protocol of Carruthers *et al.* (2018). The reads were trimmed of adapter sequences using Trimmomatic v.0.35 (Bolger *et al.* 2014) under default settings and read quality was assessed in fastQc (Andrews and FastQC 2015). An initial assembly was performed using Trinity (Grabherr *et al.* 2011) using default settings. The coverage of the assembly was assessed by mapping the original reads back to the assembly using Bowtie2 (Langmead and Salzberg 2012). Redundant transcripts were removed using CD-HIT with a cluster value of 0.99 (Fu *et al.* 2012). Each transcript was converted into the single best amino acid sequence using TransDecoder (Haas and Papanicolaou 2012).

3.5 *Ks* analysis

To determine the history of WGD in each of the *Equisetum* transcriptomes, we used the FASTKs pipeline of Mckain *et al. (McKain et al. 2016)*. Each transcriptome was BLASTed against itself to identify paralogous pairs. Each pair was aligned using MUSCLE (Edgar 2004) and the *Ks* value was calculated in codeml (Yang 2007). The frequency of *Ks* values was plotted in R and the number and position of peaks in the distribution were determined using a mixture modelling approach for model-based clustering in the *mclust* package in R (Scrucca *et al.* 2016).

3.6 Gene family assignment

Orthogroups from the sampled and assembled transcriptomes were inferred using Orthofinder v.2.2.6 (Emms and Kelly 2015). The orthology search was conducted using the Diamond sequence search program. An initial filtering step was performed on all orthogroups using custom R scripts to remove all orthogroups that did not contain at least one representative of 7 (70%) of species. Remaining orthogroups were aligned using MUSCLE and uncertain regions of each alignment were trimmed using trimal (Capella-

Gutierrez *et al.* 2009). A second filtering step was implemented with custom R scripts to remove all alignments that were shorter than 200 amino acids. Phylogenetic inference was performed on each remaining orthogroup under the best-fitting model and maximum likelihood criterion in iQtree (Nguyen *et al.* 2015), with 1000 ultra-fast bootstrap replicates (Hoang *et al.* 2018).

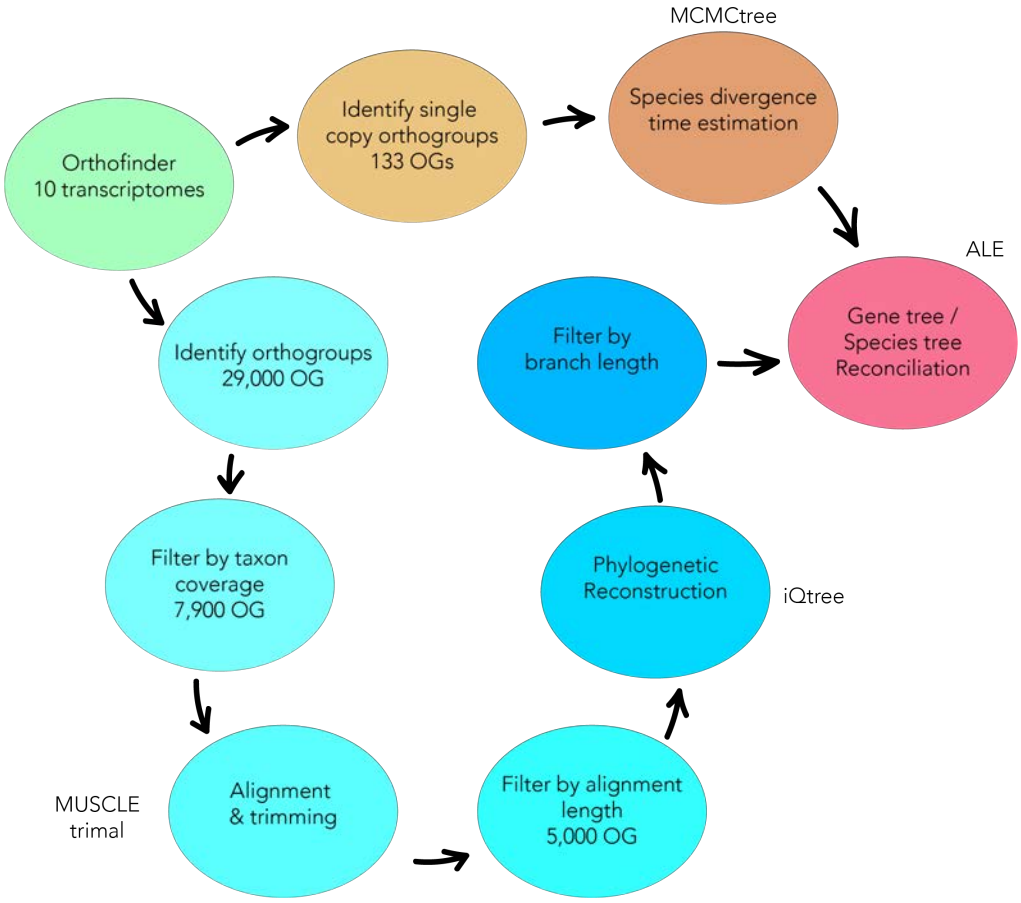


Figure 7.1. Methods used to determine gene family history from transcriptome assemblies. Specific software employed is listed next to each step, else analyses were conducted using custom R scripts.

Table 7.2. Node calibrations employed in the estimation of species divergence and genome duplication ages. The calibration on the divergence of *Hippochaete + Equisetum* was cross-calibrated in the dating of the genome duplication event.

Node	Calibration	Distribution	Ref
Tracheophyta	420.7 Ma – 451 Ma.	Uniform	Morris *et al.*
Monilophyta	384.706 Ma – 451 Ma	Uniform	Morris *et al.*
Marratiopsida	318.71 Ma – 451 Ma	Uniform	Morris *et al*
Polypodiopsida	315.1 -	Truncated Cauchy	Clark & Donoghue
Hippochaete + Equisetum	56.8 -	Truncated Cauchy	See text

3.7 Species Divergence Estimation

Single copy orthogroups were identified from the Orthofinder output and formed the basis of a two-step dating analysis. Gene families were concatenated into a superalignment and each gene was partitioned separately for a topology search performed using the edge-linked option (-spp) in iQtree (Nguyen *et al.* 2015).

The inferred topology was used to inform a fixed-topology molecular clock analysis in MCMCTREE (Yang 2007). 5 node calibrations were implemented, with a uniform distribution between minima and soft maxima (Table 7.2). Previous studies have placed the fossil taxon *Equisetum fluviatoides* as sister to *E. diffusum* (Mciver and Basinger 1989; Elgorriaga *et al.* 2018). This would provide a minimum constraint on the divergence between *E. diffusum* and *E. arvense*. However, the results of the total evidence analysis supported a placement of *E. fluviatoides* as sister to both *E. diffusum* and *E. arvense* and so we established a calibration for the divergence of the two subgenera (see below). As with the total evidence analysis, the clock rate prior was determined based on the tree length from the topology search, with a total of 0.12 substitutions site^{-1} year^{-1}. Fixing the shape parameter to two, we adjusted the scale parameter to 16 (dos Reis *et al.* 2016; Morris *et al.* 2018). Using the normal approximation method, we ran two independent analyses, each for 5,000,000 generations,

discarding the first 2,000,000 generations as burn-in. Convergence of each run was assessed using Tracer (Rambaut *et al.* 2014).

Clade Calibrated: crown-*Equisetum*: 64.96 - 451 Ma

Fossil taxon and specimen: Equisetum fluviatoides [US1-103] from 34 metres above the base of the Ravenscrag Formation at Ravenscrag Butte, 17 Km southwest of Eastend, Saskatchewan, Canada (McIver and Basinger, 1989).

Phylogenetic justification: *Equisetum fluviatoides* is identified as belonging to the crown group *Equisetum* based on vegetative and reproductive similarities to the extant species *E. fluviatile* (McIver and Basinger, 1989). A phylogenetic analysis by Elgorriaga *et al.* (2018) placed *E. fluviatoides* as sister to *E. fluviatile* or *E. diffusum* within the subgenus *Equisetum*. The present study analysed the same morphological and molecular data in a Bayesian framework and was not able to resolve the position of *E. fluviatoides* within the subgenus *Equisetum* but in a more basal position as sister to clade containing *E. fluviatile, E. diffusum* and *E. arvense*. Following our results, *E. fluviatoides* was used to calibrate the divergence of subgenus *Equisetum* from subgenus *Hippochaete*, here representing the crown group of *Equisetum*.

Minimum age: 64.96 Ma

Soft maximum age: 451

Age justification: The minimum age constraint is based on the holotype of *Equisetum fluviatoides,* described from 34 metres above the base of the Ravenscrag Formation at Ravenscrag Butte, Saskatchewan, Canada by McIver and Basinger (1989) who attribute an early Palaeocene age on the basis that the Ravenscrag Formation conformably overlies the Ferris/ No. 1 Coal Zone which at this site is established to approximate the Cretaceous-Palaeocene Boundary based on palynostratigraphy, about 3 metres below the top of magnetostratigraphic zone 29r (Lerbekmo 1985).Thus, the horizon from which the holotype of *Equisetum fluviatoides* was derived falls within the 29n

magnetozone, the minimum age of which can be constrained by the 29n-28r boundary, which is dated to 64.96 Ma (Vandenberghe *et al.* 2012).

The soft maximum constraint follows Clarke *et al.* (2011) and Morris *et al.* (2018), and the maximum age established for Tracheophyta, Euphyllophyta and Monilophyta based on the first appearance of trilete spores in the Qusaiba-1 core from the Qasim Formation, northern Saudi Arabia (Steemans *et al.* 2009). This is a Late Ordovician (Katian) formation and the oldest spores within the core co-occur with Armoricochitina nigerica, within the *Fungochitina spinifera* Biozone (= *F. fungiformis* (Paris *et al.* 2007). The base of the *F. spinifera* Biozone falls within the *Dicranograptus clingani* Biozone (*Dicellograptus morrisi* sub-biozone) (Vandenbroucke *et al.* 2008), the base of which is estimated to be 451 Ma (Cooper *et al.* 2012).

3.8 Gene tree and species tree reconciliation

Gene trees inferred from Orthofinder were reconciled against the dated species tree (Fig 7.1). Gene trees were inferred under a DTL (Duplication, Transfer, Loss) model using a maximum likelihood criterion in ALE (Amalgamated Likelihood Estimation) (Szollosi *et al.* 2012). The reconciliations were performed using 1000 ultrafast bootstrap replicates. As there is no prior hypothesis regarding an ancient hybridization or allopolyploidy in *Equisetum,* we set a low prior rate of transfer to 0.1. The total number of duplications was summed for each branch in the phylogeny.

3.9 Dating a whole genome duplication

Gene families that were inferred to have duplicated along the branch leading to *Equisetum* were sampled from the ALE output. Families that contained at least one duplication along this branch and were present within the *Ks* peak (value 0.5-1.1) were selected for a molecular clock analysis. Following the methods of Clark and Donoghue (2017), gene families were used if they a) had a clear topological signal of the WGD event b) had a topology congruent with current understanding of tracheophyte phylogeny and c) did not have a signal of

additional duplication events within *Equisetum.* Gene families were concatenated with each gene in *Equisetum* being randomly sorted to either side of the duplication event. The same set of fossil calibrations were employed as in the species divergence estimation, with the exception that the calibration within *Equisetum* was cross-calibrated on both sides of the duplication. Analyses were performed as for the species divergence estimation.

3.10 Genome size analysis

Genome size estimates (1C-values) were downloaded from the C-value database for extant taxa included in the study (Bennett and Leitch 2012). The 1C-values were estimated for fossil taxa by Franks *et al.*, who derived a linear regression model for the relationship between 1C-value and stomata guard cell length. They estimated 1C-value for members of Sphenophyllales (*Sphenophyllum)* and Calamitaceae (*Calamocladus*) as well as *Equisetum haukeanum*. Though sampling a single taxon is far from ideal, for this analysis we took the values for Sphenophyllales and Calamitaceae to be representative of each lineage. We used the linear model ($y = 1.83x - 5.46$) to convert the logged guard cell widths of other fossil *Equisetum* and to a logged 1C-value (Gould 1968; Stanich *et al.* 2009; Channing *et al.* 2011; Franks *et al.* 2012; Elgorriaga *et al.* 2015). In total 21 1C-values were obtained, and the tree was pruned to those taxa with 1C-values (Table 7.1). 1C-values were analysed as continuous characters in BayesTraits v.3 (Pagel 1999), and the ancestral 1C-value was estimated for *Equisetum* subg. *Equisetum, Equisetum* subg. *Hippochaete*, crown group *Equisetum*, Equisetaceae and Equisetales.

4 Results

4.1 Total Evidence Analysis

The total evidence analysis was able to partially resolve the backbone of the Equisetales phylogeny (Fig 7.1). The monophyly of Equisetales received strong support, with Neocalamitaceae as sister to all remaining Equisetaceae, although there was only weak

support for the monophyly of Neocalamitaceae. As with Elgorriaga *et al.,* we resolve *Equisetites arenaceus* and *Spaciinodum collinsonii* as sister to the crown group *Equisetum*, though in our analysis *Spaciinodum* is the closest outgroup.

Relationships within *Equisetum* were poorly resolved. The two subgenera *Equisetum* and *Hippochaete* were both well supported, as was the position of *E. clarnoi* and *E. fluviatoides* within each respectively. The relationships of the outgroups were also poorly resolved, including the order of divergence of the Archaeocalamitaceae and Calamitaceae, although as with previous studies we confirm that Equisetaceae did not originate from within Calamitaceae (Elgorriaga *et al.* 2018).

4.2 Divergence time estimates

The total evidence analysis proposed a Devonian origin of both Sphenopsids and ferns that is congruent with previous studies (Morris *et al.* 2018). The Sphenophyllales and Equisetales diverged during the Carboniferous along with most of the extinct lineages of Equisetales including the Archaeocalamitaceae and Calamitaceae (366-359 Ma). The divergence between Equisetaceae and *Neocalamites* occurred during the Permian (314 – 263 Ma).

It confirms the hypothesis that *Equisetum* is more ancient than molecular data alone suggest (Stanich *et al.* 2009), with crown group *Equisetum* shown to originate during the Middle Triassic – Late Jurassic (239 – 185 Ma). Both subgenera were also found to originate earlier than molecular evidence suggested, with *Equisetum* originating 142-68 Ma and *Hippochaete* with a slightly younger origin at 124 – 46 Ma.

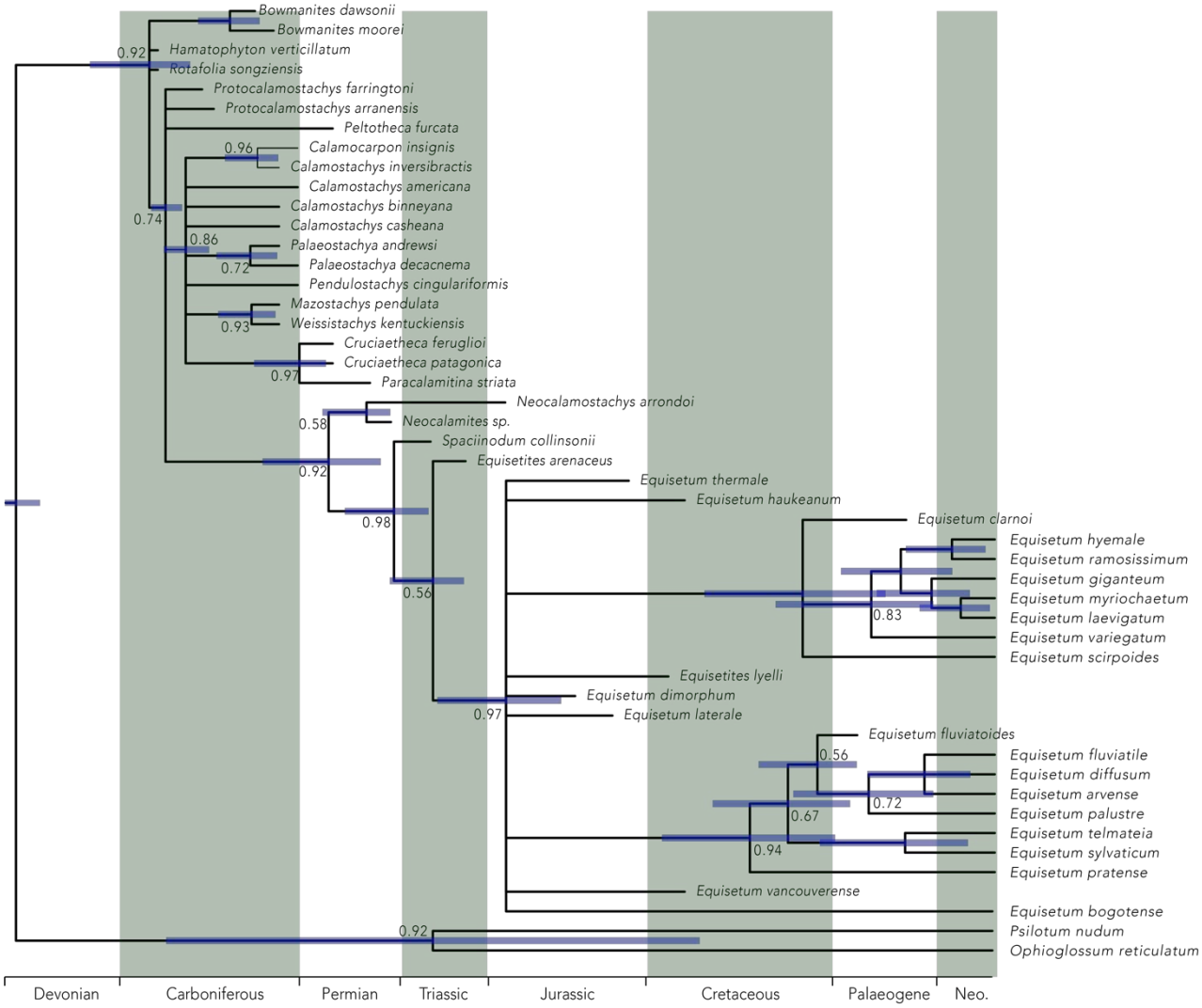


Figure 7.2. Total evidence phylogeny of extinct and extant Equisetales. The tree was constructed using Bayesian analysis of morphological and molecular data. The tree was calibrated using the ages of the fossils as tip calibrations and a uniform calibration between 451-385 Ma was applied to the root. Blue bars represent the 95% HPD interval for each node. Posterior probabilities less than 1.0 are shown next to each node.

4.3 Rates of Morphological Evolution

Rates of morphological evolution varied across the tree (Fig 7.3), with the fastest rate leading to the outgroup. The origin of major lineages was marked by the fastest rates of morphological evolution, including the Equisetales, crown group Equisetaceae and *Equisetum* subg. *Equisetum* (Fig 7.3). Generally, morphological evolution was much greater between lineages than within them, with slow rates observed within Equiseteceae and most lineages within Calamitaceae, except the branch leading to *Cruciaetheca*.

High rates of morphological evolution corresponded to large distances in the morphospace (Fig 7.4a). Most major lineages clustered tightly within the morphospace across both axes, though on the individual axes there was considerable overlap between taxa. The first axis separates Equisetaceae from Calamitaceae, while the second axis distinguishes the Equisetales from early diverging outgroups. The proportion of total disparity represented by extant taxa is low (Fig 7.4b), and disparity through time analyses showed that present levels of disparity are a small fraction of peak during the Carboniferous (Fig 7.4c). Mean disparity, measured as the average pairwise distance between taxa, within Equisetaceae (value) is lower than within Calamitaceae (value), but they do occupy a novel region of morphospace.

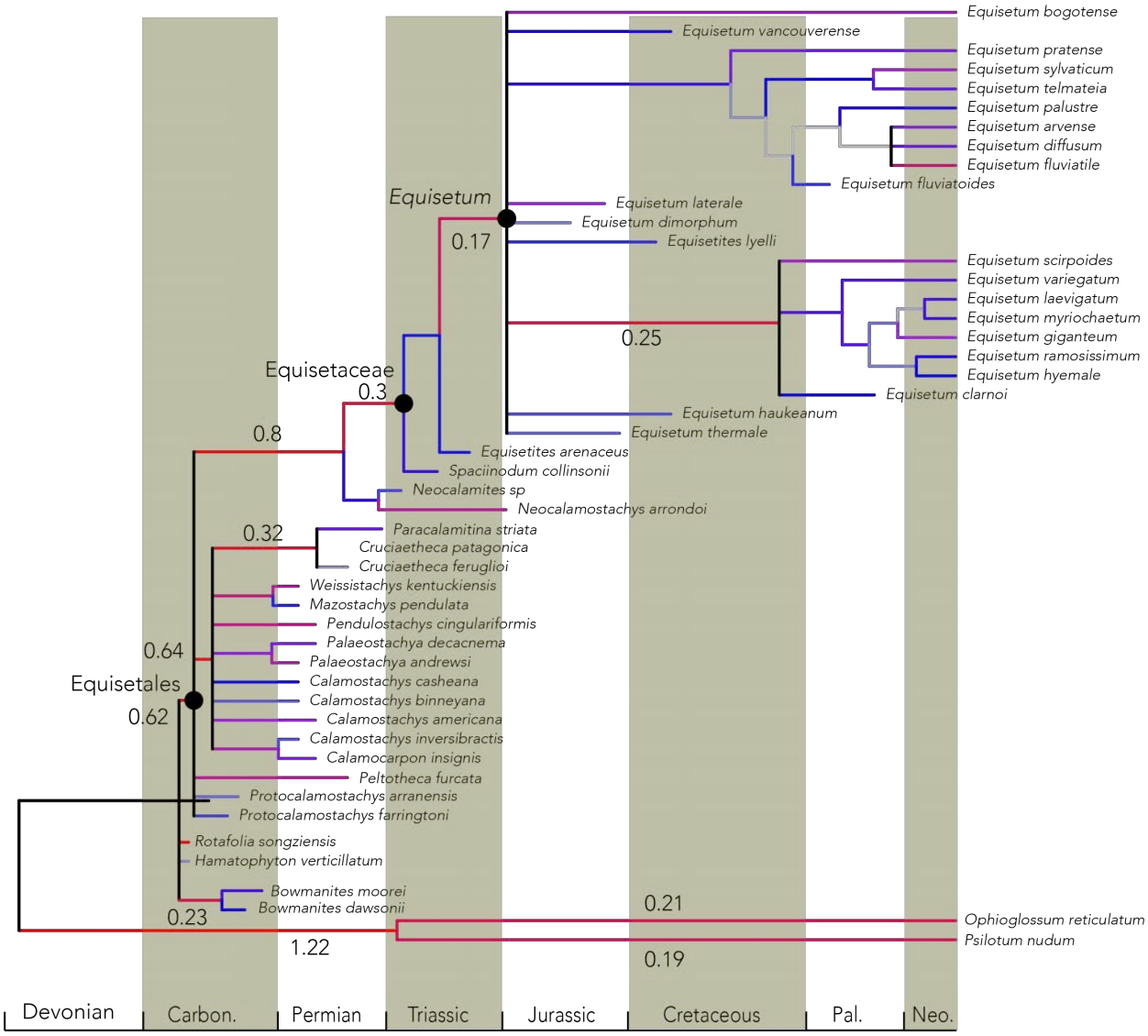


Figure 7.3. Rates of morphological evolution across the Equisetales. The branch lengths were estimated using a sample of 1000 trees from a morphological clock analysis. The branch lengths correspond to the number of character states changes per million years. High rates are shown next to the branch, with red branches representing fast rates and blue branches slow rates. Black dots represent the origin of clades of particular interest.

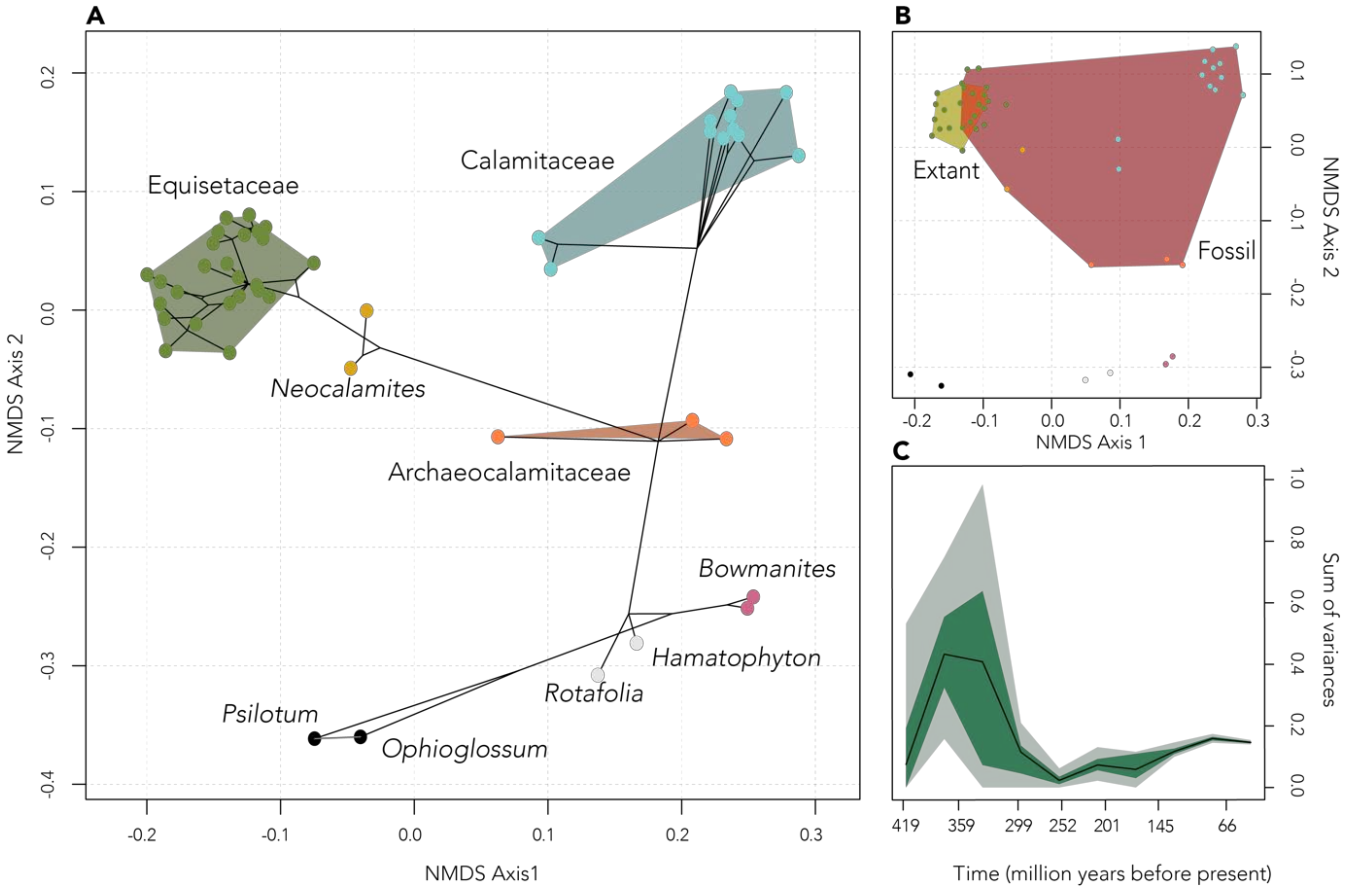


Figure 7.4. Phenotypic evolution within the Equisetales. A) An empirical phylomorphospace showing the distribution of disparity within the order. The distances between taxa were calculated using Gower's index and ordinated using non-metric multidimensional scaling (NMDS). Character states for all ancestral nodes were reconstructed and were projected into the morphospace with the tree. Convex hulls were fitted around each lineage. Colours correspond to different lineages. B) The comparative morphospace occupation of extant and fossil Equisetales. C) The evolution of disparity (sum of variances) through time estimated from the distance matrix.

4.4 Transcriptomic Analyses

The assembly of the *Equisetum arvense* transcriptome resulted in 86% of the original reads mapping back to the initial assembly, indicating reasonable coverage of reads. After clustering, 24,187 transcripts were recovered. Analysis of *Ks* values in all 4 transcriptomes revealed at least 2 peaks: one close to 0.1 representing recent duplicates, and another with a mean between 0.5 and 0.7 (Fig 7.5). The coincidence of these peaks suggests that the WGD event initially identified in *E. giganteum* is shared between all four taxa and the two subgenera. In both *E. giganteum* and *E. arvense* a small third peak was observed, but large *Ks* values become increasingly unreliable to interpret (Vanneste *et al.* 2013).

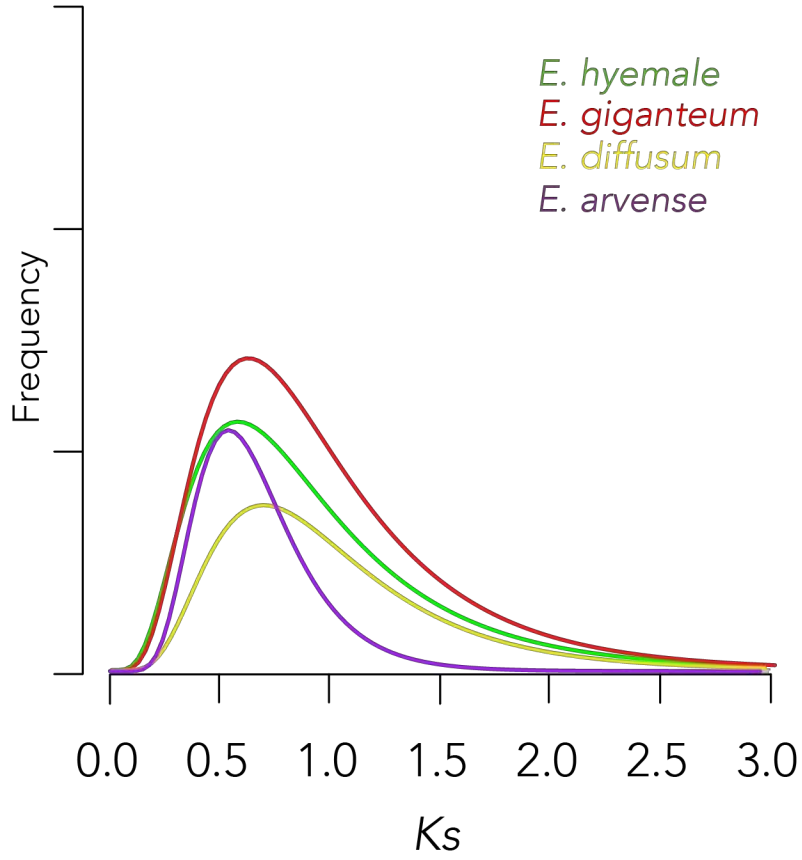


Figure 7.5. Frequency of rates of synonymous substitution (*Ks*) between paralogous pairs for 4 species of *Equisetum*. Peaks among the distributions were fitted using mixture modelling. This plot summarises the second peak in all 4 taxa corresponding to a shared ancient WGD event. Colours correspond to different taxa. The height (frequency) of each peak has been standardised.

The Orthofinder analysis initially produced 27,000 orthogroups. Two rounds of filtering based on taxon coverage and alignment length resulted in 5,009 orthogroups, each containing 70% of species and trimmed alignments > 200 amino acids. 133 single copy genes were identified for the two-step dating analysis, forming an alignment of 45671 amino acids. The age inferred for the divergence between the two subgenera was 66 – 59 Ma, much more recent compared to the total evidence analysis and with much finer confidence intervals.

The ALE analysis showed rates of duplication were generally higher on terminal branches, likely due to recent local duplication events and some of the long branches included in the study. Among all branches however, it provided strong support for a duplication event on the branch leading to *Equisetum* (Fig 7.6). 60 gene families were selected from the ALE output that showed a clear signal of the duplication event, forming an alignment of 21498 amino acids. The dating analysis estimated that the WGD event occurred 278-247 Ma, close to the Permian-Triassic boundary (Fig 7.7). When mapped onto the total evidence phylogeny, it showed that the WGD event most likely occurred either on the branch leading to Equisetaceae or following the divergence of *Spaciinodum collinsii* (Fig 7.8).

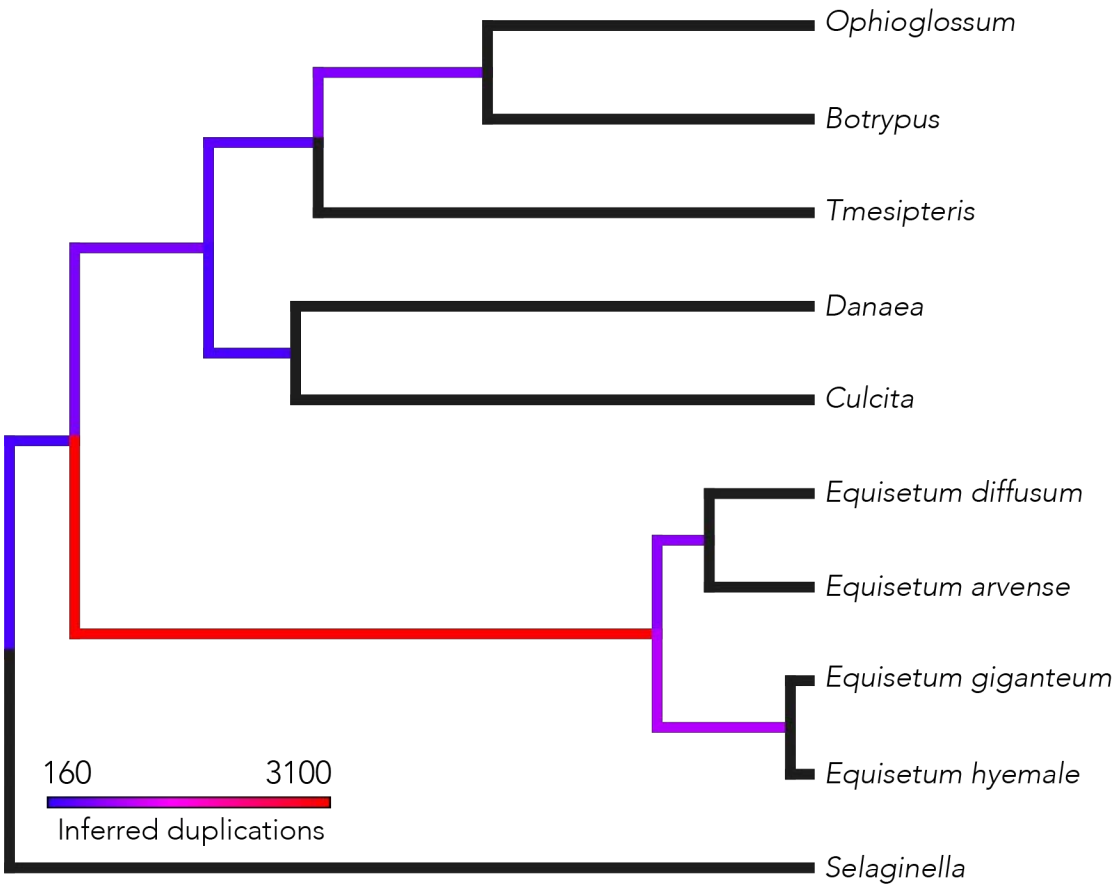


Figure 7.6. Frequency of duplications among gene families. 5,009 gene trees (including 1000 bootstrap replicates) were reconciled against the species tree. Red corresponds to a high frequency of duplication and blue to a low. Terminal branches were not visualised. The high frequency of duplications leading to *Equisetum* supports the presence of whole genome duplication.

4.5 Genome Duplication and Genome Size

Reconstruction of ancestral genome size within Sphenopsida revealed that the largest genome sizes are found within the extant *Equisetum* (ancestral 1C-value = 11.1pg), in particular the subgenus *Hippochaete* (ancestral 1C-value = 20.9pg; Fig 7.9). Across nodes we observed two large increases in genome size: from the base of *Equisetum* to *Hippochaete* (1C = 11.1 – 20.9pg) and from the base of Equisetales to Equisetaceae (1C = 3.2 - 11.1pg; Fig 7.9).

5 Discussion

5.1 The evolutionary history of Equisetales

The evolutionary history of Equisetales was reconstructed using a combination of molecular and morphological data in combined in a Bayesian 'total evidence' framework (Fig 7.2). Broadly, the relationships resolved were highly congruent with the parsimony based results of Elgorriaga *et al.* (2018) though the resolution of species relationships was more equivocal. However, it has been convincingly argued that precision achieved through parsimony is misleading and a less resolved but more accurate result is preferable (Puttick *et al.* 2017; O'Reilly *et al.* 2018). It is also possible that the exclusion of 11 continuous morphological characters from the matrix may have contributed to the loss of resolution in the present study. Despite this, our results still supported the distinction between the Calamitaceae and Equisetaceae and the hypothesis that both lineages evolved independently since the Carboniferous (Fig 7.2).

The incorporation of molecular data allowed good resolution among extant taxa and greater precision in the estimation of the evolutionary timescale of Equisetales. We confirm a Triassic-Jurassic origin for crown group *Equisetum*, a Permian-Triassic origin of Equisetaceae and a Carboniferous origin of Equisetales (Fig 7.2). Molecular estimates for the divergence between the two subgenera were considerably younger, supporting a Cretaceous divergence (Fig 7.7). However, this incongruence can be explained by either the inability of the total evidence analysis to resolve the *Equisetum* crown group, or the inability of the molecular analysis to accommodate the information provided by more fossil taxa.

5.2 Duplication and Evolution in *Equisetum*

Our analyses supported the view that there has been a single WGD event within the *Equisetum* lineage (Vanneste *et al.* 2015). Thus remains the question as to how they evolved such high chromosome numbers (Clark *et al.* 2016). Further sampling from the recent 1KP dataset may illuminate whether high chromosome numbers are ancestral within

Euphyllophyta or whether the single ancient WGD in *Equisetum* followed by conservative DNA retention has driven the increase in chromosome number. A bimodal distribution of genome sizes in modern horsetails also remains a fascinating question (Clark *et al.* 2016), since our analyses did not recover evidence of a more recent duplication event in either subgenus.

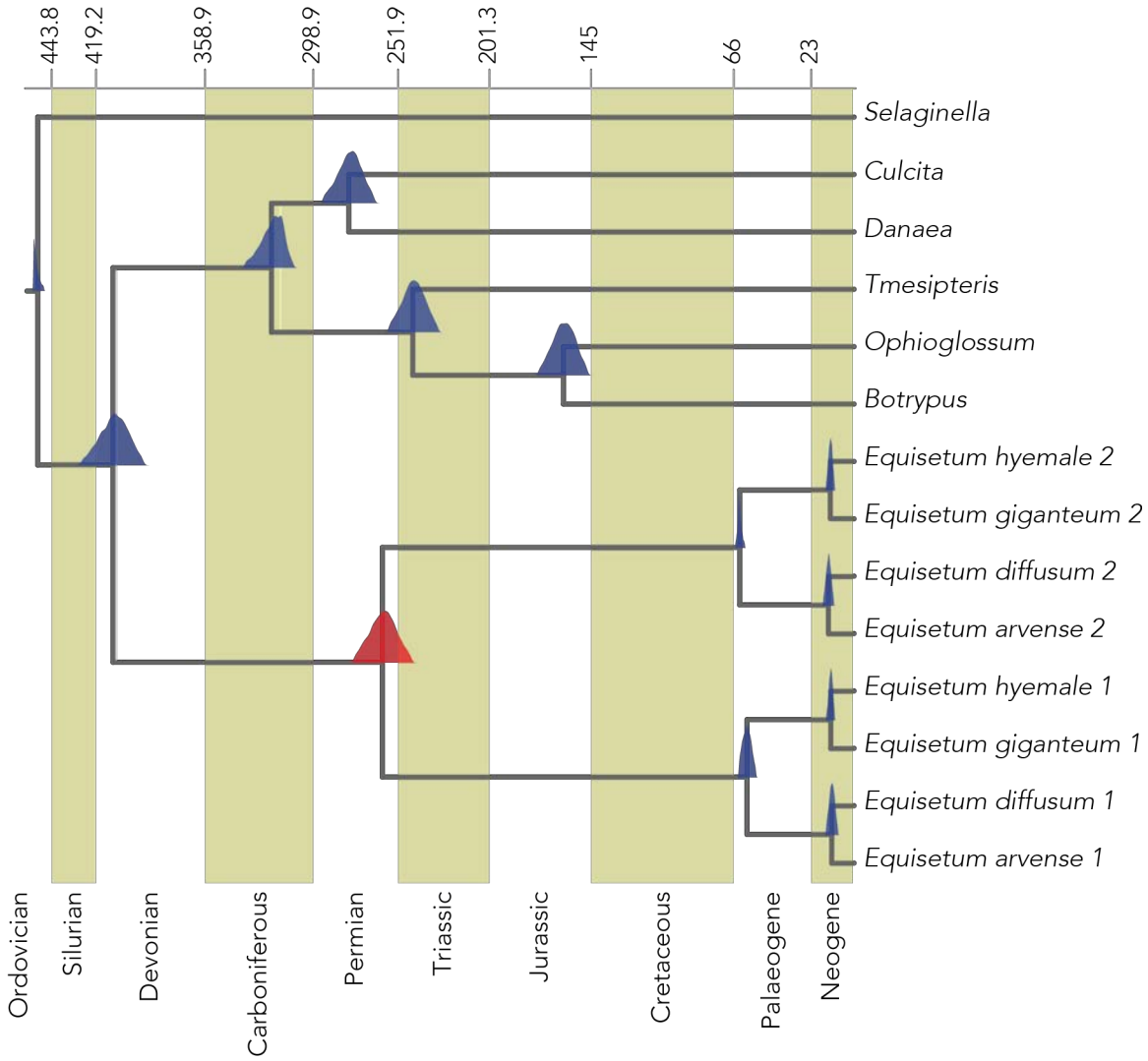


Figure 7.7. Inferred age of the whole genome duplication (WGD) event in *Equisetum*. 60 multi-copy gene families were concatenated to inform a molecular clock analysis. The 95% HPD is shown for each speciation node in blue, with the duplication node in red.

WGD has been proposed as one means to explain the distribution of phenotypic variance within the plant kingdom. There exist multiple models and a few examples demonstrating how novel traits have arisen in the wake of WGD that have been maintained and diversified on a macroevolutionary scale (Edger *et al.* 2015). However, fundamental to studying the association between WGD and macroevolution is the ability to locate the event phylogenetically and geologically (Clark and Donoghue 2018). The WGD event proposed in *E. giganteum* was known only from a single transcriptome and the geological age was poorly constrained (Vanneste *et al.* 2015). Ages inferred directly from *Ks* distributions can be inaccurate due to sequence saturation and the assumption of a strict clock (Doyle and Egan 2010; Vanneste *et al.* 2013; Clark and Donoghue 2018). By increasing taxonomic sampling, we have better resolved the position of the WGD event and find that occurred in the ancestor of both subgenera of *Equisetum* (Fig 7.5). Further, using phylogenomic methods and molecular clocks, we found that the WGD event is among the most ancient recorded in plants, occurring close to the Permian-Triassic boundary (Fig 7.7). As such, we can say that the WGD event in *Equisetum* is not associated with the K-Pg boundary and the discrepancy between the ages recovered here and those of Vanneste *et al.* (2015) suggests that other WGD events placed along this boundary should be re-examined (Clark and Donoghue 2018). The estimate achieved here is comparable in precision to recent estimates other ancient WGD in plants (Clark and Donoghue 2017) and serves to highlight the power of these methods to constrain the timing of the event to within 30 million years along one of the most isolated branches within land plants.

Importantly, these precise estimates allow us to locate the WGD event within close proximity of the origin of Equisetaceae. To elucidate a macroevolutionary role for whole genome duplication in land plant evolution, it is clearly insufficient to consider only extant taxa. *Equisetum* is known to be a poor representative of the species and morphological diversity that once existed within the Sphenopsida. As such, the inferences that can be made from studying only extant taxa are severely limited. Here, we have shown that a combination

of palaeontological and genomic approaches is the most appropriate when considering ancient or 'palaeo'-polyploidy.

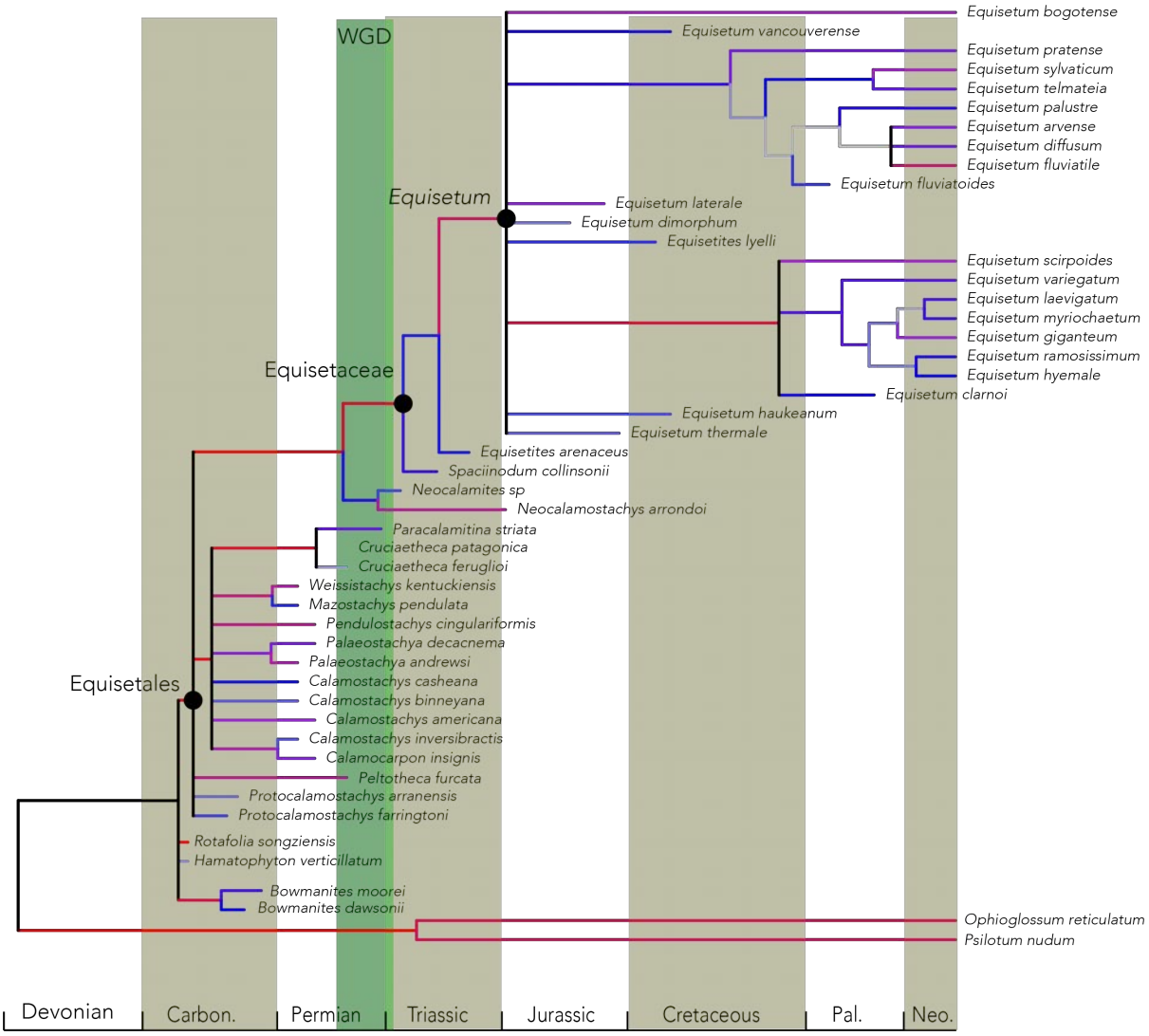


Figure 7.8. Whole genome duplication and rates of morphological evolution. The inferred age of the WGD event is shown in relation to the rates of morphological evolution. The WGD event (green) predates the evolution of Equisetaceae and corresponds with an elevated rate of morphological evolution. Branches are coloured to show high (red) and low (blue) rates of morphological evolution.

As well as the timing of the event, testing for an association between macroevolutionary change and WGD requires the quantification of any proposed outcomes (Clark and Donoghue 2018). Morphospace analysis provides a means and framework to analyse the distribution of phenotypes. Extant *Equisetum* and the fossil taxa that descended from the WGD event represent only a fraction of the morphological diversity of Equisetales (Fig 7.4b). However, by examining the rates of morphological evolution across the tree, it is possible that the WGD event coincides with a high rate (Fig 7.8). Furthermore, though not highly disparate, the evolution of Equisetaceae and the WGD event also correlates with a movement into a novel area of morphospace (Fig 7.4a).

While the WGD event does coincide with a high rate of morphological evolution, rates elsewhere along the tree are higher. Notably, the origin of Equisetales, Calamitaceae and the branch leading to *Neocalamites* + Equisetaceae all experienced faster rates (Fig 3) and in the case of Calamitaceae, greater disparity. The evolution of Equisetales is generally associated with relative stability and few character state changes. Innovations that potential coincide with the WGD event include arete spores, leaf sheath fusion and an appressed leaf tip orientation, the absence of secondary xylem (Stanich *et al.* 2009), and possibly the expression of all three reproductive regulatory modules (Tomescu *et al.* 2017). However, much of the morphology that makes *Equisetum* so distinctive, including the Equisetostele, spore elaters and whorled single-veined leaves all evolved at nodes either before or after the WGD event (Elgorriaga *et al.* 2018). Throughout the evolutionary history of Equisetales the accumulation and transformation of characters associated with the extant taxa is gradual, and many of the distinguishing features, including a compacted strobilus and small size, have evolved slowly and mosaic over several nodes (Stewart and Rothwell 1993; Taylor *et al.* 2009). This suggests that while WGD may have had a role in promoting the diversity of the Equisetaceae, it is not a prerequisite to the evolution of disparity within Equisetales. Further, given the hypothesis that WGD may have promoted the longevity of the *Equisetum* lineage (Vanneste *et al.* 2015), it is noted that of the species descended from the WGD event a

greater number have gone extinct than remain to the present day and so any advantages conferred by WGD did not aid the longevity of these lineages.

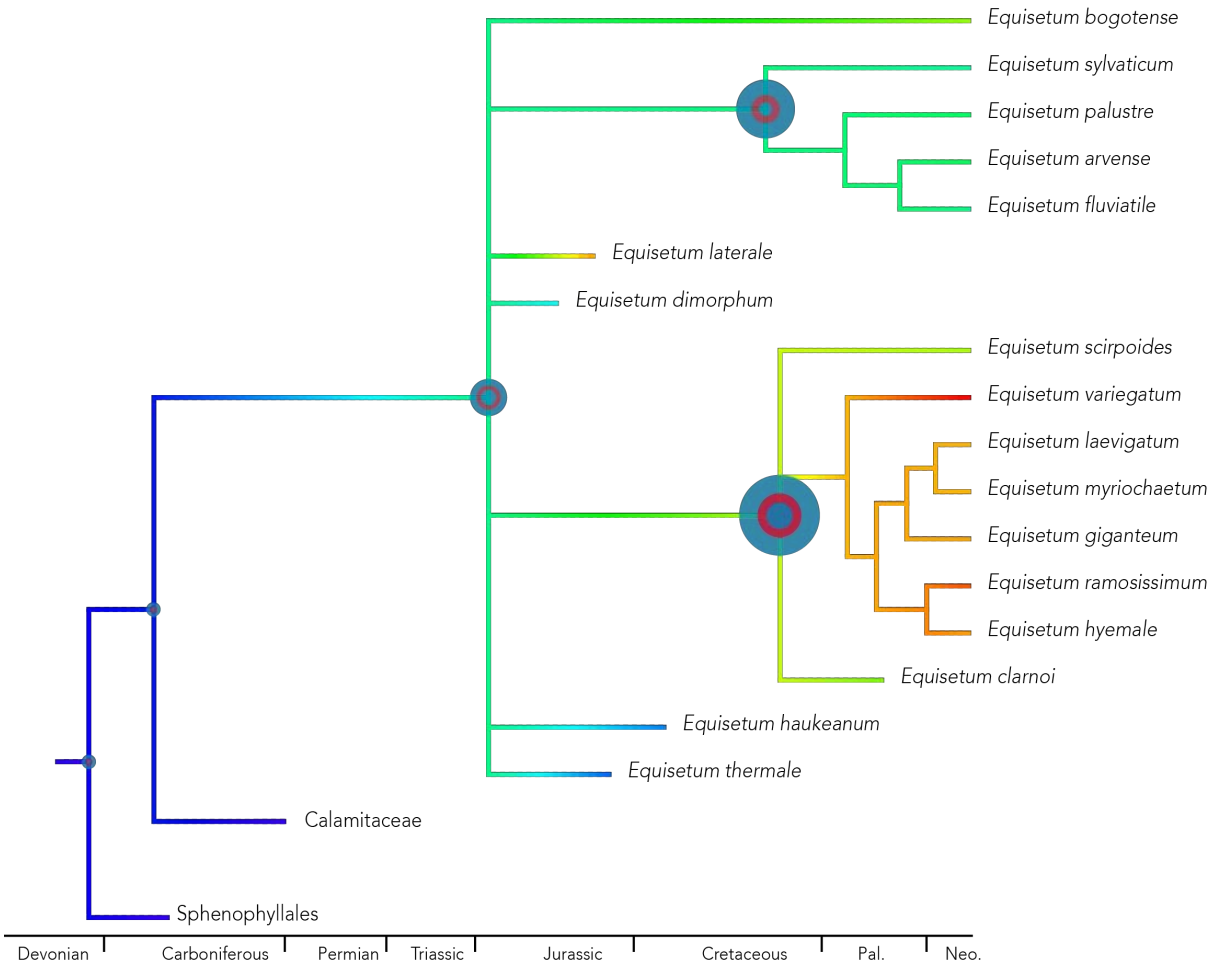


Figure 7.9. The reconstruction of ancestral genome size across the Equisetales. The genome size was reconstructed based on both extant and fossil 1C-value estimates based on stomatal width. The reconstructed size is shown at each node, with the width of the circle proportional to the 1C-value. The middle circle represents the mean estimate, while the small and large circles represent the lower and upper 95% HPD values, respectively. Branches are coloured to show the evolution of large (red) and small (blue) genome sizes.

Genome size evolution within Equisetales shows that the inferred WGD event is also associated with an increase in ancestral genome size (Fig 7.9). This is in some ways surprising, since the signal of genome duplication in genome size estimates rapidly erodes due to the dynamism of plant genomes (Leitch and Bennett 2004; Puttick *et al.* 2015). As there are no extant members of Calamitaceae, it is not possible to rule out that they may have undergone their own independent WGD event. However, the small genome size inferred for Calamitaceae (Franks *et al.* 2012) and relative stasis of fern genome evolution means that we may speculate that there may have been no WGD events in this lineage. The burst of ancient and taxonomically important new taxa discovered in recent years lends hope that further discoveries fossil taxa with exceptional preservation may shed more light on the genome evolution of this ancient lineage.

7 Conclusions

This study provides comprehensive evidence to refute the hypothesis that the WGD event during the evolution of *Equisetum* coincided with the K-Pg boundary. Instead, we find that it was far more ancient, occurring closer to the End Permian. The difference between the estimates presented here and previous studies is striking and the coincidence of several other WGD events with the K-Pg boundary certainly warrants further review. In the absence of high species diversity or 'extinction resistance', we examined morphological evolution in the wake of WGD, yet found little support for the evolution of key innovations associated with WGD. Further studies across other clades are needed to determine whether WGD and macroevolution are causally linked or whether they are simply 'ships that pass in the night'. Importantly, we demonstrate the need to incorporate palaeontological methods into the study of palaeopolyploidy, showing that a perspective which encapsulates only living taxa will often fail to detect consequences of ploidy change in plant evolution.

Chapter 8

Conclusions

James W. Clark

1 Advancements

The prevalence of whole genome duplication among land plants is undeniable, yet its importance remains controversial. This thesis presented the following aims: to improve our understanding of the relative and absolute timing of WGD events across the land plant phylogeny and to characterise the outcomes.

The relative timing of WGD events will undoubtedly be further clarified with the incoming deluge of sequence data. At the time of writing, the last remaining major branch of the plant phylogeny without a sequenced genome, the ferns, have received two. Hidden within the small genomes of two highly enigmatic, aquatic and heterosporous ferns is the signal of multiple rounds of ancient polyploidy (Li *et al.* 2018a). The 10,000 plant genome project will undoubtedly eventually uncover more WGD events (Twyford 2018), but as pointed out in chapters 1 and 4, these data need careful consideration. Identifying WGD events is the first step towards characterising their evolutionary outcomes. *Ks* plots are unable of recovering more ancient WGD events, and so as our genomic sampling looks to more early-divergent lineages we shall have to rely on phylogenomic approaches. In Chapter 4, the point is made that gene tree reconciliation methods are only as good as their underlying gene trees, and that gene tree error and uncertainty should be accounted for. These methods are highly sensitive to their inputs: inaccuracies in gene tree topology will likely infer a large number of duplications towards the root of the tree as the algorithm struggles to reconcile the large amount of topological incongruence by instead suggesting an (improbable) scenario of early duplication and extensive loss. My implementation of the Amalgamated Likelihood Estimation (ALE) software goes some way to approaching this

problem as 1) ALE can integrate over topological uncertainty by reconciling a distribution of gene tree topologies and 2) ALE allows for gene transfer between lineages, that could reconcile topological error without resorting to ancient duplications but can also be used to infer hybridisation between lineages

The second aspect of characterising WGD events is to locate them in absolute time. Estimates of age produced by *Ks* plots have a tendency to resolve ages close to the K-Pg boundary (Vanneste *et al.* 2014b; Vanneste *et al.* 2014a), or in some cases prior to the evolution of land plants (Guan *et al.* 2016; Roodt *et al.* 2017). More accurate estimation of the timing of WGD events can be derived from the gene family history (Clark and Donoghue 2017). These methods also allow co-estimation of both the age of the duplication event and the subsequent timing of species divergence. This is especially important when considering macroevolutionary outcomes of WGD, since it provides a direct measure of the 'lag' observed between cause and outcome. Applied to the two most ancient WGD events identified within plants, where previous estimates lack precision (Jiao *et al.* 2011), I was able to constrain the timing to within 30 million years for each event. The long lag between the duplication and later species divergence suggests that any macroevolutionary consequences of either event took more than 50 million years to appear.

The availability of topologically congruent gene families containing a clear signal of each WGD event decreased with older events, however I was still able to produce highly precise estimates of multiple recent and ancient duplication events. These analyses also present a familiar problem associated with any molecular clock study: our modelling of the fossil record. Molecular clocks require a careful appraisal of the fossil record and appropriate modelling of fossil constraints in order to be accurate. Reassuringly, I found in most cases the estimates to be robust to the choice of calibrations and the data best fit by more conservative calibration approaches.

Throughout this thesis, I found that the link between WGD and macroevolution is at best equivocal. In Chapter 4 I focus on diversification, where grasses show the highest rates of diversification among all monocots (and one of the highest rates among all angiosperms),

yet the increase in diversification is decoupled from the WGD event by 10-20 million years. This is corroborated by other recent studies, which have found across all angiosperms that WGD does not appear to correlate with diversification (Landis *et al.* 2018). The role of WGD in promoting morphological complexity has been more difficult to test, previously limited to either correlations between cell types and WGD (Lang *et al.* 2010) or the evolution of single traits (Edger *et al.* 2015). In Chapter 3 I assembled a dataset that captured much of the morphological variation within the plant kingdom. This dataset was used initially to answer fundamental questions about the distribution of phenotypes across the kingdom and how they evolved over geological time. The matrix also formed the basis of an analysis of morphological evolution in angiosperms following genome duplication. The angiosperm-specific WGD, *epsilon*, has been cited as the cause of the evolution of the flower (Chanderbali *et al.* 2017). Taking this example, we showed that the relationship between WGD and morphological evolution to be complex. *Epsilon* demonstrably contributed to the distinctiveness of angiosperms and variation therein, but even without the contribution of WGD angiosperms remain highly distinct and diverse. A survey of other angiosperms and *Equisetum* uncovers a similar relationship to that of WGD and diversification: sometimes WGD coincides with a burst of morphological evolution, though sometimes it does not.

2 Future Directions

As previously mentioned, the clearest next step in the field is the thorough sampling of genomic data across the land plant phylogeny. Following this however, lies a greater challenge in estimating the *significance* of WGD in plant evolution. My conclusions are that WGD only sometimes seems to produce a macroevolutionary outcome. The question that needs to be addressed is why some lineages have diversified, or evolved innovations, in the wake of WGD whereas others have not. What is the evolutionary context to make WGD successful, is it dependent on the environment and available niche, the genomic context, or is it a 'spandrel' related to clonal reproduction (Freeling 2017)?

The dating methods demonstrated in this thesis provide a means to locate WGD events in geological time and the next step should be to develop a framework where all known WGD events are dated through similarly rigorous methods. The methods used to establish evolutionary timescales are themselves evolving, and we are now able to incorporate more information from the fossils themselves into dating analyses (Ronquist *et al.* 2012; Heath *et al.* 2014). The accuracy of these methods is still uncertain (O'Reilly *et al.* 2015), but in systems where a history of WGD is complemented by an informative fossil record (Yu *et al.* 2017), they could be used to well-calibrated estimates.

Determining the role of WGD in the evolution of the land plant bodyplan will be contingent on two things: the expansion of evo-devo beyond established angiosperm systems and the continued assembly of phenotypic datasets. Examples of individual genes and gene networks evolving post-WGD has largely been limited to the Brassicales or Poaceae, yet to learn why some lineages have been successful following WGD and others haven't, we need a better understanding of which traits have arisen as a result of WGD.

To determine how great a contribution WGD has made to plant morphological evolution will rely on the generation and reassessment of more phenotypic datasets. A positive recent trend has led to researchers exploring the phenotypic evolution of multiple lineages (Chartier *et al.* 2014; Chartier *et al.* 2017; Coudert *et al.* 2017). Similar approaches based on traits that have arisen through WGD can quantify their impact on total disparity or trait-dependent diversification.

It was one of the aims of this thesis to show that a consideration of the fossil record will refine our hypotheses regarding the role of WGD in macroevolution. When considering both morphological evolution and diversification, it is crucial that extinct diversity is also considered. The morphological distinctiveness of *Equisetum* is an artefact of extinction, and it may transpire that the unique character combinations present in angiosperms evolved piecemeal in fossils yet to be discovered. To answer the most fundamental questions about WGD, it will be crucial that the advances in sequencing and bioinformatics are explored in the context of development and macroevolution.

References

Airoldi, C.A., and B. Davies. 2012. 'Gene duplication and the evolution of plant MADS-box transcription factors', *Journal of Genetics and Genomics*, 39: 157-65.

Alfaro, M.E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D.L. Rabosky, G. Carnevale, and L.J. Harmon. 2009. 'Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates', *Proceedings of the National Academy of Sciences USA*, 106: 13410-4.

Allendorf, F., and G. Thorgaard. 1984. 'Tetraploidy and the evolution of salmonid fishes.' in BJ Turner (ed.), *Evolutionary Genetics of Fishes* (New York: Plenum Press).

Andrews, S., and A. FastQC. 2015. 'A quality control tool for high throughput sequence data'. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arrigo, N., and M.S. Barker. 2012. 'Rarely successful polyploids and their legacy in plant genomes', *Current Opinion in Plant Biology*, 15: 140-6.

Banks, J.A., T. Nishiyama, M. Hasebe, J.L. Bowman, M. Gribskov, C. dePamphilis, V.A. Albert, N. Aono, T. Aoyama, B.A. Ambrose, N.W. Ashton, M.J. Axtell, E. Barker, M.S. Barker, J.L. Bennetzen, N.D. Bonawitz, C. Chapple, C. Cheng, L.G. Correa, M. Dacre, J. DeBarry, I. Dreyer, M. Elias, E.M. Engstrom, M. Estelle, L. Feng, C. Finet, S.K. Floyd, W.B. Frommer, T. Fujita, L. Gramzow, M. Gutensohn, J. Harholt, M. Hattori, A. Heyl, T. Hirai, Y. Hiwatashi, M. Ishikawa, M. Iwata, K.G. Karol, B. Koehler, U. Kolukisaoglu, M. Kubo, T. Kurata, S. Lalonde, K. Li, Y. Li, A. Litt, E. Lyons, G. Manning, T. Maruyama, T.P. Michael, K. Mikami, S. Miyazaki, S. Morinaga, T. Murata, B. Mueller-Roeber, D.R. Nelson, M. Obara, Y. Oguri, R.G. Olmstead, N. Onodera, B.L. Petersen, B. Pils, M. Prigge, S.A. Rensing, D.M. Riano-Pachon, A.W. Roberts, Y. Sato, H.V. Scheller, B. Schulz, C. Schulz, E.V. Shakirov, N. Shibagaki, N. Shinohara, D.E. Shippen, I. Sorensen, R. Sotooka, N. Sugimoto, M. Sugita, N. Sumikawa, M. Tanurdzic, G. Theissen, P. Ulvskov, S. Wakazuki, J.K. Weng, W.W. Willats, D. Wipf, P.G. Wolf, L. Yang, A.D. Zimmer, Q. Zhu, T. Mitros, U. Hellsten, D. Loque, R. Otillar, A. Salamov, J. Schmutz, H. Shapiro, E. Lindquist, S. Lucas, D. Rokhsar, and I.V. Grigoriev. 2011. 'The Selaginella genome identifies genetic changes associated with the evolution of vascular plants', *Science*, 332: 960-3.

Bansal, M.S., M. Kellis, M. Kordi, and S. Kundu. 2018. 'RANGER-DTL 2.0: Rigorous reconstruction of gene-family evolution by duplication, transfer, and loss', *Bioinformatics*.

Bapst, D.W. 2012. 'paleotree: an R package for paleontological and phylogenetic analyses of evolution', *Methods in Ecology and Evolution*, 3: 803-07.

Barba-Montoya, J., M. Dos Reis, H. Schneider, P.C.J. Donoghue, and Z. Yang. 2018. 'Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution', *New Phytologist*, 218: 819-34.

Barker, M.S., Z. Li, T.I. Kidder, C.R. Reardon, Z. Lai, L.O. Oliveira, M. Scascitelli, and L.H. Rieseberg. 2016. 'Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae', *American Journal of Botany*, 103: 1203-11.

Bateman, R.M., P.R. Crane, W.A. DiMichele, P.R. Kenrick, N.P. Rowe, T. Speck, and W.E. Stein. 1998. 'Early evolution of land plants: Phylogeny, physiology, and ecology of the primary terrestrial radiation', *Annual Review of Ecology and Systematics*, 29: 263-92.

Beer, C., M. Reichstein, E. Tomelleri, P. Ciais, M. Jung, N. Carvalhais, C. Rodenbeck, M.A. Arain, D. Baldocchi, G.B. Bonan, A. Bondeau, A. Cescatti, G. Lasslop, A. Lindroth, M. Lomas, S. Luyssaert, H. Margolis, K.W. Oleson, O. Roupsard, E. Veenendaal, N. Viovy, C. Williams, F.I. Woodward, and D. Papale. 2010. 'Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate', *Science*, 329: 834-8.

Bennett, M., and I. Leitch. 2012. "Plant DNA C-values database." In.: Royal Botanic Gardens Kew.

Birchler, J.A., and R.A. Veitia. 2012. 'Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines', *Proceedings of the National Academy of Sciences USA*, 109: 14746-53.

Blanc, G., and K.H. Wolfe. 2004. 'Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes', *Plant Cell*, 16: 1667-78.

Blomberg, S.P., T. Garland, Jr., and A.R. Ives. 2003. 'Testing for phylogenetic signal in comparative data: behavioral traits are more labile', *Evolution*, 57: 717-45.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30: 2114-20.

Bond, W.J., G.F. Midgley, and F.I. Woodward. 2003. 'The importance of low atmospheric CO2 and fire in promoting the spread of grasslands and savannas', *Global Change Biology*, 9: 973-82.

Bonnefille, R. 2010. 'Cenozoic vegetation, climate changes and hominid evolution in tropical Africa', *Global and Planetary Change*, 72: 390-411.

Bottani, S., N.R. Zabet, J.F. Wendel, and R.A. Veitia. 2018. 'Gene expression dominance in allopolyploids: Hypotheses and models', *Trends in Plant Science*, 23: 393-402.

Bouchenak-Khelladi, Y., A.M. Muasya, and H.P. Linder. 2014. 'A revised evolutionary history of Poales: origins and diversification', *Botanical Journal of the Linnean Society*, 175: 4-16.

Boyce, C.K. 2008. 'How green was Cooksonia? The importance of size in understanding the early evolution of physiology in the vascular plant lineage', *Paleobiology*, 34: 179-94.

Boyce, C.K., and A.H. Knoll. 2002. 'Evolution of developmental potential and the multiple independent origins of leaves in Paleozoic vascular plants', *Paleobiology*, 28: 70-100.

Boyce, C.K., and A.B. Leslie. 2012. 'The paleontological context of angiosperm vegetative evolution', *International Journal of Plant Sciences*, 173: 561-68.

Cannon, S.B., M.R. McKain, A. Harkess, M.N. Nelson, S. Dash, M.K. Deyholos, Y. Peng, B. Joyce, C.N. Stewart, Jr., M. Rolf, T. Kutchan, X. Tan, C. Chen, Y. Zhang, E. Carpenter, G.K. Wong, J.J. Doyle, and J. Leebens-Mack. 2015. 'Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes', *Molecular Biology and Evolution*, 32: 193-210.

Capella-Gutierrez, S., J.M. Silla-Martinez, and T. Gabaldon. 2009. 'trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses', *Bioinformatics*, 25: 1972-3.

Carruthers, M., A.A. Yurchenko, J.J. Augley, C.E. Adams, P. Herzyk, and K.R. Elmer. 2018. 'De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species', *BMC Genomics*, 19: 32.

Catarino, B., A.J. Hetherington, D.M. Emms, S. Kelly, and L. Dolan. 2016. 'The stepwise increase in the number of transcription factor families in the precambrian predated the diversification of plants on land', *Molecular Biology and Evolution*, 33: 2815-19.

Chanderbali, A.S., B.A. Berger, D.G. Howarth, P.S. Soltis, and D.E. Soltis. 2016. 'Evolving ideas on the origin and evolution of flowers: New perspectives in the genomic era', *Genetics*, 202: 1255-65.

Chanderbali, A.S., B.A. Berger, D.G. Howarth, D.E. Soltis, and P.S. Soltis. 2017. 'Evolution of floral diversity: genomics, genes and *gamma*', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372.

Channing, A., A. Zamuner, D. Edwards, and D. Guido. 2011. 'Equisetum thermale sp. nov. (Equisetales) from the Jurassic San Agustin hot spring deposit, Patagonia: anatomy, paleoecology, and inferred paleoecophysiology', *American Journal of Botany*, 98: 680-97.

Chartier, M., S. Lofstrand, M. von Balthazar, S. Gerber, F. Jabbour, H. Sauquet, and J. Schonenberger. 2017. 'How (much) do flowers vary? Unbalanced disparity among flower functional modules and a mosaic pattern of morphospace occupation in the order Ericales', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20170066.

Chartier, M., F. Jabbour, S. Gerber, P. Mitteroecker, H. Sauquet, M. von Balthazar, Y. Staedler, P.R. Crane, and J. Schoenenberger. 2014. 'The floral morphospace - a modern comparative approach to study angiosperm evolution', *New Phytologist*, 204: 841-53.

Chen, C.-Y., S.-T. Chen, H.-F. Juan, and H.-C. Huang. 2012. 'Lengthening of 3′UTR increases with morphological complexity in animal evolution', *Bioinformatics*, 28: 3178-81.

Chomicki, G., M. Coiro, and S.S. Renner. 2017. 'Evolution and ecology of plant architecture: integrating insights from the fossil record, extant morphology, developmental genetics and phylogenies', *Annals of Botany*, 120: 855-91.

Christin, P.A., E. Spriggs, C.P. Osborne, C.A. Stromberg, N. Salamin, and E.J. Edwards. 2014. 'Molecular dating, evolutionary rates, and the age of the grasses', *Systematic Biology*, 63: 153-65.

Christin, P.A., G. Besnard, E. Samaritani, M.R. Duvall, T.R. Hodkinson, V. Savolainen, and N. Salamin. 2008. 'Oligocene CO2 decline promoted C4 photosynthesis in grasses', *Current Biology*, 18: 37-43.

Clark, J., O. Hidalgo, J. Pellicer, H. Liu, J. Marquardt, Y. Robert, M. Christenhusz, S. Zhang, M. Gibby, I.J. Leitch, and H. Schneider. 2016. 'Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny', *New Phytologist*, 210: 1072-82.

Clark, J.W., and P.C.J. Donoghue. 2017. 'Constraining the timing of whole genome duplication in plant evolutionary history', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20170912.

———. 2018. 'Whole-genome duplication and plant macroevolution', *Trends in Plant Science*, 23: 933-45.

Clarke, J., P.C.J. Donoghue, and R.C.M. Warnock. 2011. 'Establishing a timescale for plant evolution', *New Phytologist*, 192: 266-301.

Clarke, J.T., G.T. Lloyd, and M. Friedman. 2016. 'Little evidence for enhanced phenotypic evolution in early teleosts relative to their living fossil sister group', *Proceedings of the National Academy of Sciences USA*, 113: 11531-36.

Coen, E.S., and E.M. Meyerowitz. 1991. 'The war of the whorls: genetic interactions controlling flower development', *Nature*, 353: 31-7.

Coiro, M., G. Chomicki, and J.A. Doyle. 2018. 'Experimental signal dissection and method sensitivity analyses reaffirm the potential of fossils and morphology in the resolution of the relationship of angiosperms and Gnetales', *Paleobiology*, 44: 490-510.

Comparot-Moss, S., and K. Denyer. 2009. 'The evolution of the starch biosynthetic pathway in cereals and other grasses', *Journal of Experimental Botany*, 60: 2481-92.

Conant, G.C., and K.H. Wolfe. 2008. 'Turning a hobby into a job: how duplicated genes find new functions', *Nature Reviews Genetics*, 9: 938-50.

Conant, G.C., J.A. Birchler, and J.C. Pires. 2014. 'Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time', *Current Opinion in Plant Biology*, 19: 91-8.

Cooper, R., P. Sadler, O. Hammer, and F. Gradstein. 2012. 'The Ordovician Period.' in, *The Geologic Time Scale* (Elsevier).

Coudert, Y., N.E. Bell, C. Edelin, and C.J. Harrison. 2017. 'Multiple innovations underpinned branching form diversification in mosses', *New Phytologist*, 215: 840-50.

Crow, K.D., and G.P. Wagner. 2006. 'What is the role of genome duplication in the evolution of complexity and diversity?', *Molecular Biology and Evolution*, 23: 887-92.

Cui, L., P.K. Wall, J.H. Leebens-Mack, B.G. Lindsay, D.E. Soltis, J.J. Doyle, P.S. Soltis, J.E. Carlson, K. Arumuganathan, A. Barakat, V.A. Albert, H. Ma, and C.W. dePamphilis. 2006. 'Widespread genome duplications throughout the history of flowering plants', *Genome Research*, 16: 738-49.

D'Hont, A., F. Denoeud, J.M. Aury, F.C. Baurens, F. Carreel, O. Garsmeur, B. Noel, S. Bocs, G. Droc, M. Rouard, C. Da Silva, K. Jabbari, C. Cardi, J. Poulain, M. Souquet, K. Labadie, C. Jourda, J. Lengelle, M. Rodier-Goud, A. Alberti, M. Bernard, M. Correa, S. Ayyampalayam, M.R. McKain, J. Leebens-Mack, D. Burgess, M. Freeling, A.M.D. Mbeguie, M. Chabannes, T. Wicker, O. Panaud, J. Barbosa, E. Hribova, P. Heslop-Harrison, R. Habas, R. Rivallan, P. Francois, C. Poiron, A. Kilian, D. Burthia, C. Jenny, F. Bakry, S. Brown, V. Guignon, G. Kema, M. Dita, C. Waalwijk, S. Joseph, A. Dievart, O. Jaillon, J. Leclercq, X. Argout, E. Lyons, A. Almeida, M. Jeridi, J. Dolezel, N. Roux, A.M. Risterucci, J. Weissenbach, M. Ruiz, J.C. Glaszmann, F. Quetier, N. Yahiaoui, and P. Wincker. 2012. 'The banana (Musa acuminata) genome and the evolution of monocotyledonous plants', *Nature*, 488: 213-7.

De La Torre, A.R., Z. Li, Y. Van de Peer, and P.K. Ingvarsson. 2017. 'Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants', *Molecular Biology and Evolution*, 34: 1363-77.

de Vries, J., and J.M. Archibald. 2018. 'Plant evolution: landmarks on the path to terrestrial life', *New Phytologist*, 217: 1428-34.

Deline, B. 2009. 'The effects of rarity and abundance distributions on measurements of local morphological disparity', *Paleobiology*, 35: 175-89.

Deline, B., J.M. Greenwood, J.W. Clark, M.N. Puttick, K.J. Peterson, and P.C.J. Donoghue. 2018. 'Evolution of metazoan morphological disparity', *Proceedings of the National Academy of Sciences USA*, 115: 8909-18.

Des Marais, David L., Alan R. Smith, Donald M. Britton, and Kathleen M. Pryer. 2003. 'Phylogenetic relationships and evolution of extant horsetails, *Equisetum*, based on chloroplast DNA sequence data (*rbcL* and *trnL-F*)', *International Journal of Plant Sciences*, 164: 737-51.

Devos, N., P. Szovenyi, D.J. Weston, C.J. Rothfels, M.G. Johnson, and A.J. Shaw. 2016. 'Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta)', *New Phytologist*, 211: 300-18.

Dixon, P. 2003. 'VEGAN, a package of R functions for community ecology', *Journal of Vegetation Science*, 14: 927-30.

Dodsworth, S., M.W. Chase, and A.R. Leitch. 2016. 'Is post-polyploidization diploidization the key to the evolutionary success of angiosperms?', *Botanical Journal of the Linnean Society*, 180: 1-5.

Donoghue, P.C., and M.J. Benton. 2007. 'Rocks and clocks: calibrating the Tree of Life using fossils and molecules', *Trends Ecol Evol*, 22: 424-31.

Donoghue, P.C., and Z. Yang. 2016. 'The evolution of methods for establishing evolutionary timescales', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371.

Donoghue, P.C.J., and M.A. Purnell. 2005. 'Genome duplication, extinction and vertebrate evolution', *Trends in Ecology & Evolution*, 20: 312-19.

dos Reis, M., P.C. Donoghue, and Z. Yang. 2016. 'Bayesian molecular clock dating of species divergences in the genomics era', *Nature Reviews Genetics*, 17: 71-80.

Doyle, J.A. 2008. 'Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower', *International Journal of Plant Sciences*, 169: 816-43.

Doyle, J.A. 2012. 'Phylogenetic analyses and morphological innovations in land plants.' in, *Annual Plant Reviews Volume 45* (John Wiley & Sons, Ltd.).

Doyle, J.A., and P.K. Endress. 2010. 'Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots', *Journal of Systematics and Evolution*, 48: 1-35.

Doyle, J.J., and A.N. Egan. 2010. 'Dating the origins of polyploidy events', *New Phytologist*, 186: 73-85.

Edgar, R.C. 2004. 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32: 1792-7.

Edger, P.P., J.C. Hall, A. Harkess, M. Tang, J. Coombs, S. Mohammadin, M.E. Schranz, Z. Xiong, J. Leebens-Mack, B.C. Meyers, K.J. Sytsma, M.A. Koch, I.A. Al-Shehbaz, and J.C. Pires. 2018. 'Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses', *American Journal of Botany*, 105: 463-69.

Edger, P.P., H.M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glockner, A.E. Platts, D.G. Heckel, J.P. Der, E.K. Wafula, M. Tang, J.A. Hofberger, A. Smithson, J.C. Hall, M. Blanchette, T.E. Bureau, S.I. Wright, C.W. dePamphilis, M. Eric Schranz, M.S. Barker, G.C. Conant, N. Wahlberg, H. Vogel, J.C. Pires, and C.W. Wheat. 2015. 'The butterfly plant arms-race escalated by gene and genome duplications', *Proceedings of the National Academy of Sciences USA*, 112: 8362-6.

Edwards, E.J., C.P. Osborne, C.A. Stromberg, S.A. Smith, C.G. Consortium, W.J. Bond, P.A. Christin, A.B. Cousins, M.R. Duvall, D.L. Fox, R.P. Freckleton, O. Ghannoum, J. Hartwell, Y. Huang, C.M. Janis, J.E. Keeley, E.A. Kellogg, A.K. Knapp, A.D. Leakey, D.M. Nelson, J.M. Saarela, R.F. Sage, O.E. Sala, N. Salamin, C.J. Still, and B. Tipple. 2010. 'The origins of C4 grasslands: integrating evolutionary and ecosystem science', *Science*, 328: 587-91.

Elgorriaga, A., I.H. Escapa, B. Bomfleur, R. Cuneo, and E.G. Ottone. 2015. 'Reconstruction and phylogenetic significance of a new *Equisetum* Linnaeus species from the Lower Jurassic of Cerro Bayo (Chubut Province, Argentina)', *Ameghiniana*, 52: 135-52.

Elgorriaga, A., I.H. Escapa, G.W. Rothwell, A.M.F. Tomescu, and N. Ruben Cuneo. 2018. 'Origin of *Equisetum*: Evolution of horsetails (Equisetales) within the major euphyllophyte clade Sphenopsida', *American Journal of Botany*, 105: 1286-303.

Emms, D.M., and S. Kelly. 2015. 'OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy', *Genome Biology*, 16: 157.

Erwin, D.H. 2007. 'Disparity: Morphological pattern and developmental context', *Palaeontology*, 50: 57-73.

Estep, M.C., M.R. McKain, D. Vela Diaz, J. Zhong, J.G. Hodge, T.R. Hodkinson, D.J. Layton, S.T. Malcomber, R. Pasquet, and E.A. Kellogg. 2014. 'Allopolyploidy, diversification, and the Miocene grassland expansion', *Proceedings of the National Academy of Sciences USA* 111: 15149-54.

Fawcett, J.A., and Y. Van de Peer. 2010. 'Angiosperm polyploids and their road to evolutionary success', *2010*, 2.

Fawcett, J.A., S. Maere, and Y. Van de Peer. 2009. 'Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event', *Proceedings of the National Academy of Sciences USA* 106: 5737-42.

Fernandez-Mazuecos, M., and B.J. Glover. 2017. 'The evo-devo of plant speciation', *Nature Ecology and Evolution*, 1: 110.

Floyd, S.K., and J.L. Bowman. 2007. 'The ancestral developmental tool kit of land plants', *International Journal of Plant Sciences*, 168: 1-35.

Foote, M. 1993. 'Contributions of Individual Taxa to Overall Morphological Disparity', *Paleobiology*, 19: 403-19.

Foster, C.S.P., H. Sauquet, M. van der Merwe, H. McPherson, M. Rossetto, and S.Y.W. Ho. 2017. 'Evaluating the Impact of Genomic Data and Priors on Bayesian Estimates of the Angiosperm Evolutionary Timescale', *Systematic Biology*, 66: 338-51.

Franks, P.J., R.P. Freckleton, J.M. Beaulieu, I.J. Leitch, and D.J. Beerling. 2012. 'Megacycles of atmospheric carbon dioxide concentration correlate with fossil plant genome size', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367: 556-64.

Freeling, M. 2009. 'Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition', *Annu Rev Plant Biol*, 60: 433-53.

Freeling, M. 2017. 'The distribution of ancient polyploidies in the plant phylogenetic tree is a spandrel of occasional sex', *The Plant Cell*.

Freeling, M., and B.C. Thomas. 2006. 'Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity', *Genome Res*, 16.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. 'CD-HIT: accelerated for clustering the next-generation sequencing data', *Bioinformatics*, 28: 3150-2.

Fujisawa, M., Y. Shima, H. Nakagawa, M. Kitagawa, J. Kimbara, T. Nakano, T. Kasumi, and Y. Ito. 2014. 'Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins', *Plant Cell*, 26: 89-101.

Galtier, N., and V. Daubin. 2008. 'Dealing with incongruence in phylogenomic analyses', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363: 4023-9.

Garsmeur, O., J.C. Schnable, A. Almeida, C. Jourda, A. D'Hont, and M. Freeling. 2014. 'Two evolutionarily distinct classes of paleopolyploidy', *Molecular Biology and Evolution*, 31: 448-54.

Givnish, T.J., M. Ames, J.R. McNeal, M.R. McKain, P.R. Steele, C.W. dePamphilis, S.W. Graham, J.C. Pires, D.W. Stevenson, W.B. Zomlefer, B.G. Briggs, M.R. Duvall, M.J. Moore, J.M. Heaney, D.E. Soltis, P.S. Soltis, K. Thiele, and J.H. Leebens-Mack. 2010. 'Assembling the Tree of the Monocotyledons: Plastome Sequence Phylogeny and Evolution of Poales', *Annals of the Missouri Botanical Garden*, 97: 584-616.

Glennon, K.L., M.E. Ritchie, and K.A. Segraves. 2014. 'Evidence for shared broad-scale climatic niches of diploid and polyploid plants', *Ecology Letters*, 17: 574-82.

Goloboff, P.A., A. Torres, and J.S. Arias. 2018. 'Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology', *Cladistics*, 34: 407-37.

Gould, R.E. 1968. 'Morphology of *Equisetum laterale* Phillips, 1829, and *E. bryanii* sp. nov. from the Mesozoic of South-Eastern Queensland', *Australian Journal of Botany*, 16: 153-76.

Gould, S.J. 1990. *Wonderful life: the Burgess Shale and the nature of history* (WW Norton & Company).

Gower, J.C. 1971. 'A general coefficient of similarity and some of its properties', *Biometrics*, 27: 857-71.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, and Q. Zeng. 2011. 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nat Biotechnol*, 29.

Gramzow, L., L. Weilandt, and G. Theissen. 2014. 'MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants', *Annals of Botany*, 114: 1407-29.

Gregg, W.C.T., S.H. Ather, and M.W. Hahn. 2017. 'Gene-tree reconciliation with MUL-trees to resolve polyploidy events', *Systematic Biology*, 66: 1007-18.

Gregory, T.R. 2005. 'The C-value enigma in plants and animals: a review of parallels and an appeal for partnership', *Annals of Botany*, 95: 133-46.

M.W. Chase, M.J.M. Christenhusz, M.F. Fay, J.W. Byng, W.S. Judd, D.E. Soltis, D.J. Mabberley, A.N. Sennikov, P.S. Soltis, and P.F. Stevens. 2016. 'An update of

the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV', *Botanical Journal of the Linnean Society*, 181: 1-20.

Guan, R., Y. Zhao, H. Zhang, G. Fan, X. Liu, W. Zhou, C. Shi, J. Wang, W. Liu, X. Liang, Y. Fu, K. Ma, L. Zhao, F. Zhang, Z. Lu, S.M. Lee, X. Xu, J. Wang, H. Yang, C. Fu, S. Ge, and W. Chen. 2016. 'Draft genome of the living fossil Ginkgo biloba', *Gigascience*, 5: 49.

Guillerme, T. 2015. 'dispRity: a package for measuring disparity in R.', *Zenodo*.

Guillerme, T., N. Cooper, and A. Smith. 2018. 'Time for a rethink: time sub-sampling methods in disparity-through-time analyses', *Palaeontology*, 61: 481-93.

Haas, B., and A. Papanicolaou. 2012. 'Transdecoder'. https://transdecoder.github.io/.

Hafeez, M., M. Shabbir, F. Altaf, and A.A. Abbasi. 2016. 'Phylogenomic analysis reveals ancient segmental duplications in the human genome', *Molecular Phylogenetics and Evolution*, 94: 95-100.

Hanada, K., C. Zou, M.D. Lehti-Shiu, K. Shinozaki, and S.-H. Shiu. 2008. 'Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli', *Plant Physiology*, 148: 993-1003.

Hanschen, E.R., T.N. Marriage, P.J. Ferris, T. Hamaji, A. Toyoda, A. Fujiyama, R. Neme, H. Noguchi, Y. Minakuchi, M. Suzuki, H. Kawai-Toyooka, D.R. Smith, H. Sparks, J. Anderson, R. Bakaric, V. Luria, A. Karger, M.W. Kirschner, P.M. Durand, R.E. Michod, H. Nozaki, and B.J. Olson. 2016. 'The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity', *Nature Communications*, 7: 11370.

Harrison, C.J. 2017. 'Development and genetics in the evolution of land plant body plans', *Philosophical Transactions of the Royal Society B-Biological Sciences*, 372: 20150490.

Heath, T.A., J.P. Huelsenbeck, and T. Stadler. 2014. 'The fossilized birth-death process for coherent calibration of divergence-time estimates', *Proceedings of the National Academy of Sciences USA*, 111: E2957-66.

Herendeen, P.S., E.M. Friis, K.R. Pedersen, and P.R. Crane. 2017. 'Palaeobotanical redux: revisiting the age of the angiosperms', *Nature Plants*, 3: 17015.

Hetherington, A.J., E. Sherratt, M. Ruta, M. Wilkinson, B. Deline, P.C.J. Donoghue, and K. Angielczyk. 2015. 'Do cladistic and morphometric data capture common patterns of morphological disparity?', *Palaeontology*, 58: 393-99.

Hidalgo, O., J. Pellicer, M.J.M. Christenhusz, H. Schneider, and I.J. Leitch. 2017. 'Genomic gigantism in the whisk-fern family (Psilotaceae): Tmesipteris obliqua challenges record holder Paris japonica', *Botanical Journal of the Linnean Society*, 183: 509-14.

Hilton, J., and R.M. Bateman. 2006. 'Pteridosperms are the backbone of seed-plant phylogeny', *Journal of the Torrey Botanical Society*, 133: 119-68.

Hoang, D.T., O. Chernomor, A. von Haeseler, B.Q. Minh, and L.S. Vinh. 2018. 'UFBoot2: Improving the ultrafast bootstrap approximation', *Molecular Biology and Evolution*, 35: 518-22.

Hoegg, S., H. Brinkmann, J.S. Taylor, and A. Meyer. 2004. 'Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish', *Journal of Molecular Evolution*, 59: 190-203.

Huang, C.-H., C. Zhang, M. Liu, Y. Hu, T. Gao, J. Qi, and H. Ma. 2016. 'Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics', *Molecular Biology and Evolution*, 33: 2820-35.

Huelsenbeck, J.P., R. Nielsen, and J.P. Bollback. 2003. 'Stochastic mapping of morphological characters', *Systematic Biology*, 52: 131-58.

Huerta-Cepas, J., S. Capella-Gutierrez, L.P. Pryszcz, M. Marcet-Houben, and T. Gabaldon. 2014. 'PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome', *Nucleic Acids Research*, 42: D897-902.

Hughes, M., S. Gerber, and M.A. Wills. 2013. 'Clades reach highest morphological disparity early in their evolution', *Proceedings of the National Academy of Sciences USA*, 110: 13875-9.

Humphreys, A.M., and H.P. Linder. 2013. 'Evidence for recent evolution of cold tolerance in grasses suggests current distribution is not limited by (low) temperature', *New Phytologist*, 198: 1261-73.

Iles, W.J.D., S.Y. Smith, M.A. Gandolfo, and S.W. Graham. 2015. 'Monocot fossils suitable for molecular dating analyses', *Botanical Journal of the Linnean Society*, 178: 346-74.

Inoue, J.G., P.C.J. Donoghue, and Z. Yang. 2010. 'The impact of the representation of fossil calibrations on bayesian estimation of species divergence times', *Systematic Biology*, 59: 74-89.

Jaillon, O., J.M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyere, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M.E. Pe, G. Valle, M. Morgante, M. Caboche, A.F. Adam-Blondon, J. Weissenbach, F. Quetier, P. Wincker, and C.

French-Italian Public Consortium for Grapevine Genome. 2007. 'The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla', *Nature*, 449: 463-7.

Jiao, Y., J. Li, H. Tang, and A.H. Paterson. 2014. 'Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots', *Plant Cell*, 26: 2792-802.

Jiao, Y., N.J. Wickett, S. Ayyampalayam, A.S. Chanderbali, L. Landherr, P.E. Ralph, L.P. Tomsho, Y. Hu, H. Liang, P.S. Soltis, D.E. Soltis, S.W. Clifton, S.E. Schlarbaum, S.C. Schuster, H. Ma, J. Leebens-Mack, and C.W. dePamphilis. 2011. 'Ancestral polyploidy in seed plants and angiosperms', *Nature*, 473: 97-100.

Jiao, Y., J. Leebens-Mack, S. Ayyampalayam, J.E. Bowers, M.R. McKain, J. McNeal, M. Rolf, D.R. Ruzicka, E. Wafula, N.J. Wickett, X. Wu, Y. Zhang, J. Wang, Y. Zhang, E.J. Carpenter, M.K. Deyholos, T.M. Kutchan, A.S. Chanderbali, P.S. Soltis, D.W. Stevenson, R. McCombie, J.C. Pires, G.K. Wong, D.E. Soltis, and C.W. Depamphilis. 2012. 'A genome triplication associated with early diversification of the core eudicots', *Genome Biology*, 13: R3.

Julca, I., M. Marcet-Houben, P. Vargas, and T. Gabaldon. 2017. 'Phylogenomics of the olive tree (Olea europaea) disentangles ancient allo- and autopolyploidizations in Lamiales', *bioRxiv*.

Kagale, S., S.J. Robinson, J. Nixon, R. Xiao, T. Huebert, J. Condie, D. Kessler, W.E. Clarke, P.P. Edger, M.G. Links, A.G. Sharpe, and I.A. Parkin. 2014. 'Polyploid evolution of the Brassicaceae during the Cenozoic era', *Plant Cell*, 26: 2777-91.

Kaplan, D.R. 2001. 'The science of plant morphology: definition, history, and role in modern biology', *American Journal of Botany*, 88: 1711-41.

Kasahara, M. 2007. 'The 2R hypothesis: an update', *Current Opinion in Immunology*, 19: 547-52.

Katoh, K., and D.M. Standley. 2013. 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability', *Molecular Biology and Evolution*, 30: 772-80.

Kellogg, E.A. 2000. 'The grasses: A case study in macroevolution', *Annual Review of Ecology and Systematics*, 31: 217-38.

———. 2016. 'Has the connection between polyploidy and diversification actually been tested?', *Current Opinion in Plant Biology*, 30: 25-32.

Kelly, L.J., S. Renny-Byfield, J. Pellicer, J. Macas, P. Novak, P. Neumann, M.A. Lysak, P.D. Day, M. Berger, M.F. Fay, R.A. Nichols, A.R. Leitch, and I.J. Leitch. 2015. 'Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of

DNA removal characterizes extreme expansions in genome size', *New Phytologist*, 208: 596-607.

Kenrick, P., and P.R. Crane. 1997. *The origin and early diversification of land plants: a cladistic study* (Smithsonian Institution Press: Washington D.C.).

Laenen, B., B. Shaw, H. Schneider, B. Goffinet, E. Paradis, A. Desamore, J. Heinrichs, J.C. Villarreal, S.R. Gradstein, S.F. McDaniel, D.G. Long, L.L. Forrest, M.L. Hollingsworth, B. Crandall-Stotler, E.C. Davis, J. Engel, M. Von Konrat, E.D. Cooper, J. Patino, C.J. Cox, A. Vanderpoorten, and A.J. Shaw. 2014. 'Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts', *Nature Communications*, 5: 5134.

Landis, J.B., D.E. Soltis, Z. Li, H.E. Marx, M.S. Barker, D.C. Tank, and P.S. Soltis. 2018. 'Impact of whole-genome duplication events on diversification rates in angiosperms', *American Journal of Botany*, 105: 348-63.

Lang, D., and S.A. Rensing. 2015. 'The evolution of transcriptional regulation in the viridiplantae and its correlation with morphological complexity.' in, *Evolutionary transitions to multicellular life* (Springer).

Lang, D., B. Weiche, G. Timmerhaus, S. Richardt, D.M. Riano-Pachon, L.G. Correa, R. Reski, B. Mueller-Roeber, and S.A. Rensing. 2010. 'Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity', *Genome Biology and Evolution*, 2: 488-503.

Langmead, B., and S.L. Salzberg. 2012. 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9: 357-9.

Laurent, S., N. Salamin, and M. Robinson-Rechavi. 2017. 'No evidence for the radiation time lag model after whole genome duplications in Teleostei', *PLoS One*, 12: e0176384.

Leitch, I.J., and M.D. Bennett. 2004. 'Genome downsizing in polyploid plants', *Biological Journal of the Linnean Society*, 82: 651-63.

Lerbekmo, J.F. 1985. 'Magnetostratigraphic and biostratigraphic correlations of Maastrichtian to Early Paleocene strata between South-Central Alberta and Southwestern Saskatchewan', *Bulletin of Canadian Petroleum Geology*, 33: 213-26.

Leslie, A.B., and C.K. Boyce. 2012. 'Ovule function and the evolution of angiosperm reproductive innovations', *International Journal of Plant Sciences*, 173: 640-48.

Lewis, E.B. 1951. "Pseudoallelism and gene evolution." In *Cold Spring Harbor Symposia on Quantitative Biology*, 159-74. Cold Spring Harbor Laboratory Press.

Li, F.W., P. Brouwer, L. Carretero-Paulet, S. Cheng, J. de Vries, P.M. Delaux, A. Eily, N. Koppers, L.Y. Kuo, Z. Li, M. Simenc, I. Small, E. Wafula, S. Angarita, M.S.

Barker, A. Brautigam, C. dePamphilis, S. Gould, P.S. Hosmani, Y.M. Huang, B. Huettel, Y. Kato, X. Liu, S. Maere, R. McDowell, L.A. Mueller, K.G.J. Nierop, S.A. Rensing, T. Robison, C.J. Rothfels, E.M. Sigel, Y. Song, P.R. Timilsena, Y. Van de Peer, H. Wang, P.K.I. Wilhelmsson, P.G. Wolf, X. Xu, J.P. Der, H. Schluepmann, G.K. Wong, and K.M. Pryer. 2018a. 'Fern genomes elucidate land plant evolution and cyanobacterial symbioses', *Nature Plants*, 4: 460-72.

Li, M., H. An, R. Angelovici, C. Bagaza, A. Batushansky, L. Clark, V. Coneva, M. Donoghue, E. Edwards, D. Fajardo, H. Fang, M. Frank, T. Gallaher, S. Gebken, T. Hill, S. Jansky, B. Kaur, P. Klahs, L. Klein, V. Kuraparthy, J. Londo, Z. Migicovsky, A. Miller, R. Mohn, S. Myles, W. Otoni, J.C. Pires, E. Riffer, S. Schmerler, E. Spriggs, C. Topp, A. Van Deynze, K. Zhang, L. Zhu, B.M. Zink, and D.H. Chitwood. 2017. 'Persistent homology demarcates a leaf morphospace', *bioRxiv*.

Li, Z., A.E. Baniaga, E.B. Sessa, M. Scascitelli, S.W. Graham, L.H. Rieseberg, and M.S. Barker. 2015. 'Early genome duplications in conifers and other seed plants', *Science Advances*, 1: e1501084.

Li, Z., G.P. Tiley, S.R. Galuska, C.R. Reardon, T.I. Kidder, R.J. Rundell, and M.S. Barker. 2018b. 'Multiple large-scale gene and genome duplications during the evolution of hexapods', *Proceedings of the National Academy of Sciences USA*, 115: 4713-18.

Libertin, M., J. Kvacek, J. Bek, V. Zarsky, and P. Storch. 2018. 'Sporophytes of polysporangiate land plants from the early Silurian period may have been photosynthetically autonomous', *Nature Plants*, 4: 269-71.

Linder, H.P., and P.J. Rudall. 2005. 'Evolutionary history of Poales', *Annual Review of Ecology Evolution and Systematics*, 36: 107-24.

Linder, H.P., C.E.R. Lehmann, S. Archibald, C.P. Osborne, and D.M. Richardson. 2018. 'Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation', *Biological Reviews of the Cambridge Philosophical Society*, 93: 1125-44.

Lloyd, A., A. Blary, D. Charif, C. Charpentier, J. Tran, S. Balzergue, E. Delannoy, G. Rigaill, and E. Jenczewski. 2018. 'Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop', *New Phytologist*, 217: 367-77.

Lloyd, G.T. 2016. 'Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions', *Biological Journal of the Linnean Society*, 118: 131-51.

Lohaus, R., and Y. Van de Peer. 2016. 'Of dups and dinos: evolution at the K/Pg boundary', *Current Opinion in Plant Biology*, 30: 62-9.

Lynch, M., and J.S. Conery. 2000. 'The evolutionary fate and consequences of duplicate genes', *Science*, 290.

Lynch, M., and A.G. Force. 2000. 'The origin of interspecific genomic incompatibility via gene duplication', *American Naturalist*, 156: 590-605.

Lyons, E., B. Pedersen, J. Kane, M. Alam, R. Ming, H. Tang, X. Wang, J. Bowers, A. Paterson, D. Lisch, and M. Freeling. 2008. 'Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids', *Plant Physiology*, 148: 1772-81.

MacKintosh, C., and D.E.K. Ferrier. 2017. 'Recent advances in understanding the roles of whole genome duplications in evolution', *F1000Res*, 6: 1623.

Maclean, C.J., and D. Greig. 2011. 'Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast', *Evolution*, 65: 932-45.

Macqueen, D.J., and I.A. Johnston. 2014. 'A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification', *Proceedings of the Royal Society B: Biological Sciences*, 281: 20132881.

Madlung, A. 2013. 'Polyploidy and its effect on evolutionary success: old questions revisited with new tools', *Heredity (Edinb)*, 110: 99-104.

Marcet-Houben, M., and T. Gabaldon. 2015. 'Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage', *PLoS Biology*, 13: e1002220.

Martin, K.J., and P.W. Holland. 2014. 'Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication', *Molecular Biology and Evolution*, 31: 2592-611.

Mayrose, I., S.H. Zhan, C.J. Rothfels, N. Arrigo, M.S. Barker, L.H. Rieseberg, and S.P. Otto. 2015. 'Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis et al. (2014)', *New Phytologist*, 206: 27-35.

Mciver, E.E., and J.F. Basinger. 1989. 'The morphology and relationships of *Equisetum fluviatoides* sp.nov. from the Paleocene Ravenscrag formation of Saskatchewan, Canada', *Canadian Journal of Botany-Revue Canadienne De Botanique*, 67: 2937-43.

McKain, M.R., N. Wickett, Y. Zhang, S. Ayyampalayam, W.R. McCombie, M.W. Chase, J.C. Pires, C.W. dePamphilis, and J. Leebens-Mack. 2012. 'Phylogenomic analysis of transcriptome data elucidates co-occurrence of a

paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae)', *American Journal of Botany*, 99: 397-406.

McKain, M.R., H. Tang, J.R. McNeal, S. Ayyampalayam, J.I. Davis, C.W. dePamphilis, T.J. Givnish, J.C. Pires, D.W. Stevenson, and J.H. Leebens-Mack. 2016. 'A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales', *Genome Biology and Evolution*, 8: 1150-64.

Melzer, R., and G. Theissen. 2016. 'The significance of developmental robustness for species diversity', *Annals of Botany*, 117: 725-32.

Minelli, A. 2016. 'Species diversity vs. morphological disparity in the light of evolutionary developmental biology', *Annals of Botany*, 117: 781-94.

Minelli, A. 2018. *Plant evolutionary developmental biology: The evolvability of the phenotype* (Cambridge University Press).

Mishler, B.D., and S.P. Churchill. 1985. 'Transition to a land flora: phylogenetic relationships of the green algae and bryophytes', *Cladistics*, 1: 305-28.

Mishler, B.D., L.A. Lewis, M.A. Buchheim, K.S. Renzaglia, D.J. Garbary, C.F. Delwiche, F.W. Zechman, T.S. Kantz, and R.L. Chapman. 1994. 'Phylogenetic relationships of the green algae and bryophytes', *Annals of the Missouri Botanical Garden*, 81: 451-83.

Mitteroecker, P., and S.M. Huttegger. 2015. 'The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors', *Biological Theory*, 4: 54-67.

Morris, J.L., M.N. Puttick, J.W. Clark, D. Edwards, P. Kenrick, S. Pressel, C.H. Wellman, Z. Yang, H. Schneider, and P.C.J. Donoghue. 2018. 'The timescale of early land plant evolution', *Proceedings of the National Academy of Sciences USA*, 115: E2274-E83.

Mueller, B., C. Engard, and J.W.v. Goethe. 1952. 'Botanical writings', *Trans. B Mueller, introduction CJ Engard.(Honolulu, HI: University of Hawaii Press)*.

Muir, C.D., and M.W. Hahn. 2015. 'The limited contribution of reciprocal gene loss to increased speciation rates following whole-genome duplication', *American Naturalist*, 185: 70-86.

Murat, F., A. Armero, C. Pont, C. Klopp, and J. Salse. 2017. 'Reconstructing the genome of the most recent common ancestor of flowering plants', *Nature Genetics*, 49: 490-+.

Nakazato, T., M.K. Jung, E.A. Housworth, L.H. Rieseberg, and G.J. Gastony. 2006. 'Genetic map-based analysis of genome structure in the homosporous fern Ceratopteris richardii', *Genetics*, 173: 1585-97.

Nandi, O.I., M.W. Chase, and P.K. Endress. 1998. 'A combined cladistic analysis of angiosperms using rbcL and non-molecular data sets', *Annals of the Missouri Botanical Garden*, 85: 137-212.

Naugolnykh, S.V. 2004. 'On some aberrations of extant horsetails (*Equisetum* L.) and the origin of the family Equisetaceae', *Paleontological Journal*, 38: 335-42.

Nguyen, L.T., H.A. Schmidt, A. von Haeseler, and B.Q. Minh. 2015. 'IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Mol Biol Evol*, 32: 268-74.

O'Reilly, J.E., M. Dos Reis, and P.C.J. Donoghue. 2015. 'Dating tips for divergence-time estimation', *Trends in Genetics*, 31: 637-50.

O'Reilly, J.E., M.N. Puttick, D. Pisani, and P.C.J. Donoghue. 2018. 'Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data', *Palaeontology*, 61: 105-18.

O'Reilly, J.E., M.N. Puttick, L. Parry, A.R. Tanner, J.E. Tarver, J. Fleming, D. Pisani, and P.C.J. Donoghue. 2016. 'Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data', *Biology Letters*, 12.

Ohno, S. 1970. *Evolution by gene duplication* (Springer Science Business Media: Berlin).

Oyston, J.W., M. Hughes, S. Gerber, and M.A. Wills. 2016. 'Why should we investigate the morphological disparity of plant clades?', *Annals of Botany*, 117: 859-79.

Oyston, J.W., M. Hughes, P.J. Wagner, S. Gerber, and M.A. Wills. 2015. 'What limits the morphological disparity of clades?', *Interface Focus*, 5: 20150042.

Page, C. 1972. 'An interpretation of the morphology and evolution of the cone and shoot of Equisetum', *Botanical Journal of the Linnean Society*, 65: 359-97.

Pagel, M. 1999. 'Inferring the historical patterns of biological evolution', *Nature*, 401: 877-84.

Panchy, N., M. Lehti-Shiu, and S.H. Shiu. 2016. 'Evolution of gene duplication in plants', *Plant Physiology*, 171: 2294-316.

Parham, J.F., P.C. Donoghue, C.J. Bell, T.D. Calway, J.J. Head, P.A. Holroyd, J.G. Inoue, R.B. Irmis, W.G. Joyce, D.T. Ksepka, J.S. Patane, N.D. Smith, J.E. Tarver, M. van Tuinen, Z. Yang, K.D. Angielczyk, J.M. Greenwood, C.A. Hipsley, L. Jacobs, P.J. Makovicky, J. Muller, K.T. Smith, J.M. Theodor, R.C. Warnock, and M.J. Benton. 2012. 'Best practices for justifying fossil calibrations', *Systematic Biology*, 61: 346-59.

Paris, F., A. Le Hérissé, O. Monod, H. Kozlu, J.-F. Ghienne, W.T. Dean, M. Vecoli, and Y. Günay. 2007. 'Ordovician chitinozoans and acritarchs from southern and southeastern Turkey', *Revue de Micropaléontologie*, 50: 81-107.

Paterson, A.H., J.E. Bowers, and B.A. Chapman. 2004. 'Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics', *Proceedings of the National Academy of Sciences USA*, 101: 9903-8.

Pellicer, J., M.F. Fay, and I.J. Leitch. 2010. 'The largest eukaryotic genome of them all?', *Botanical Journal of the Linnean Society*, 164: 10-15.

Pellicer, J., O. Hidalgo, S. Dodsworth, and I.J. Leitch. 2018. 'Genome size diversity and its impact on the evolution of land plants', *Genes*, 9: E88.

Petersen, G., O. Seberg, A. Cuenca, D.W. Stevenson, M. Thadeo, J.I. Davis, S. Graham, and T.G. Ross. 2016. 'Phylogeny of the Alismatales (Monocotyledons) and the relationship of *Acorus* (Acorales)', *Cladistics*, 32: 141-59.

Pires, N.D., and L. Dolan. 2012. 'Morphological evolution in land plants: new designs with old genes', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367: 508-18.

Prasad, V., C.A. Stromberg, A.D. Leache, B. Samant, R. Patnaik, L. Tang, D.M. Mohabey, S. Ge, and A. Sahni. 2011. 'Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae', *Nature Communications*, 2: 480.

Preston, J.C., and E.A. Kellogg. 2007. 'Conservation and divergence of APETALA1/FRUITFULL-like gene function in grasses: evidence from gene expression analyses', *The Plant Journal*, 52: 69-81.

Puttick, M.N., J. Clark, and P.C.J. Donoghue. 2015. 'Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms', *Proceedings of the Royal Society B: Biological Sciences*, 282.

Puttick, M.N., J.E. O'Reilly, A.R. Tanner, J.F. Fleming, J. Clark, L. Holloway, J. Lozano-Fernandez, L.A. Parry, J.E. Tarver, D. Pisani, and P.C.J. Donoghue. 2017. 'Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data', *Proceedings of the Royal Society B-Biological Sciences*, 284.

Puttick, M.N., J.L. Morris, T.A. Williams, C.J. Cox, D. Edwards, P. Kenrick, S. Pressel, C.H. Wellman, H. Schneider, D. Pisani, and P.C.J. Donoghue. 2018. 'The interrelationships of land plants and the nature of the ancestral embryophyte', *Current Biology*, 28: 733-45 e2.

Qiao, X., H. Yin, L. Li, R. Wang, J. Wu, J. Wu, and S. Zhang. 2018. 'Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*)', *Frontiers in Plant Science*, 9: E61.

Rabosky, D.L. 2014. 'Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees', *PLoS One*, 9: e89543.

Rabosky, D.L., G.J. Slater, and M.E. Alfaro. 2012. 'Clade age and species richness are decoupled across the eukaryotic tree of life', *PLoS Biology*, 10: e1001381.

Rambaut, A., M. Suchard, and A.J. Drummond. 2014. "Tracer v1. 6." In.

Ramsey, J., and D.W. Schemske. 2002. 'Neopolyploidy in flowering plants', *Annual Review of Ecology and Systematics*, 33: 589-639.

Ran, J.H., T.T. Shen, M.M. Wang, and X.Q. Wang. 2018. 'Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms', *Proceedings of the Royal Society B: Biological Sciences*, 285.

Raup, D.M., and S.J. Gould. 1974. 'Stochastic simulation and evolution of morphology - towards a nomothetic paleontology', *Systematic Zoology*, 23: 305-22.

Ren, R., H. Wang, C. Guo, N. Zhang, L. Zeng, Y. Chen, H. Ma, and J. Qi. 2018. 'Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms', *Molecular Plant*, 11: 414-28.

Renne, P.R., C.J. Sprain, M.A. Richards, S. Self, L. Vanderkluysen, and K. Pande. 2015. 'State shift in Deccan volcanism at the Cretaceous-Paleogene boundary, possibly induced by impact', *Science*, 350: 76-78.

Renny-Byfield, S., and J.F. Wendel. 2014. 'Doubling down on genomes: polyploidy and crop plants', *American Journal of Botany*, 101: 1711-25.

Rensing, S.A. 2014. 'Gene duplication as a driver of plant morphogenetic evolution', *Current Opinion in Plant Biology*, 17: 43-8.

———. 2018. 'Plant evolution: Phylogenetic relationships between the earliest land plants', *Current Biology*, 28: R210-R13.

Revell, L.J. 2012. 'phytools: an R package for phylogenetic comparative biology (and other things)', *Methods in Ecology and Evolution*, 3: 217-23.

Robertson, F.M., M.K. Gundappa, F. Grammes, T.R. Hvidsten, A.K. Redmond, S. Lien, S.A.M. Martin, P.W.H. Holland, S.R. Sandve, and D.J. Macqueen. 2017. 'Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification', *Genome Biology*, 18: 111.

Romano, M., N. Brocklehurst, and J. Frobisch. 2017. 'Discrete and continuous character-based disparity analyses converge to the same macroevolutionary signal: a case study from captorhinids', *Scientific Reports*, 7: 17531.

Ronquist, F., and J.P. Huelsenbeck. 2003. 'MrBayes 3: Bayesian phylogenetic inference under mixed models', *Bioinformatics*, 19: 1572-4.

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D.L. Murray, and A.P. Rasnitsyn. 2012. 'A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera', *Systematic Biology*, 61: 973-99.

Roodt, D., R. Lohaus, L. Sterck, R.L. Swanepoel, Y. Van de Peer, and E. Mizrachi. 2017. 'Evidence for an ancient whole genome duplication in the cycad lineage', *PLoS One*, 12: e0184454.

Rothfels, C.J., A.K. Johnson, P.H. Hovenkamp, D.L. Swofford, H.C. Roskam, C.R. Fraser-Jenkins, M.D. Windham, and K.M. Pryer. 2015. 'Natural hybridization between genera that diverged from each other approximately 60 million years ago', *American Naturalist*, 185: 433-42.

Rothwell, G.W. 1996. 'Pteridophytic evolution: An often underappreciated phytological success story', *Review of Palaeobotany and Palynology*, 90: 209-22.

Ruiz-Trillo, I., and A.M. Nedelcu. 2015. *Evolutionary transitions to multicellular life: principles and mechanisms* (Springer).

Ruprecht, C., R. Lohaus, K. Vanneste, M. Mutwil, Z. Nikoloski, Y. Van de Peer, and S. Persson. 2017. 'Revisiting ancestral polyploidy in plants', *Science Advances*, 3: e1603195.

Salman-Minkov, A., N. Sabath, and I. Mayrose. 2016. 'Whole-genome duplication as a key factor in crop domestication', *Nature Plants*, 2: 16115.

Sauquet, H., and S. Magallon. 2018. 'Key questions and challenges in angiosperm macroevolution', *New Phytologist*, 219: 1170-87.

Schnable, J.C., M. Freeling, and E. Lyons. 2012. 'Genome-wide analysis of syntenic gene deletion in the grasses', *Genome Biology and Evolution*, 4: 265-77.

Schneider, H., A.R. Smith, and K.M. Pryer. 2009. 'Is morphology really at odds with molecules in estimating fern phylogeny?', *Systematic Botany*, 34: 455-75.

Schrager-Lavelle, A., H. Klein, A. Fisher, and M. Bartlett. 2017. 'Grass flowers: An untapped resource for floral evo-devo', *Journal of Systematics and Evolution*, 55: 525-41.

Schranz, M.E., S. Mohammadin, and P.P. Edger. 2012. 'Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model', *Current Opinion in Plant Biology*, 15: 147-53.

Schwager, E.E., P.P. Sharma, T. Clarke, D.J. Leite, T. Wierschin, M. Pechmann, Y. Akiyama-Oda, L. Esposito, J. Bechsgaard, T. Bilde, A.D. Buffry, H. Chao, H. Dinh, H. Doddapaneni, S. Dugan, C. Eibner, C.G. Extavour, P. Funch, J. Garb, L.B. Gonzalez, V.L. Gonzalez, S. Griffiths-Jones, Y. Han, C. Hayashi, M. Hilbrant, D.S.T. Hughes, R. Janssen, S.L. Lee, I. Maeso, S.C. Murali, D.M. Muzny, R. Nunes da Fonseca, C.L.B. Paese, J. Qu, M. Ronshaugen, C. Schomburg, A. Schonauer, A. Stollewerk, M. Torres-Oliva, N. Turetzek, B.

Vanthournout, J.H. Werren, C. Wolff, K.C. Worley, G. Bucher, R.A. Gibbs, J. Coddington, H. Oda, M. Stanke, N.A. Ayoub, N.M. Prpic, J.F. Flot, N. Posnien, S. Richards, and A.P. McGregor. 2017. 'The house spider genome reveals an ancient whole-genome duplication during arachnid evolution', *BMC Biology*, 15: 62.

Scrucca, L., M. Fop, T.B. Murphy, and A.E. Raftery. 2016. 'mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models', *The R Journal*, 8: 289-317.

Seoighe, C., and C. Gehring. 2004. 'Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome', *Trends Genet*, 20: 461-4.

Shen, H., D. Jin, J.P. Shu, X.L. Zhou, M. Lei, R. Wei, H. Shang, H.J. Wei, R. Zhang, L. Liu, Y.F. Gu, X.C. Zhang, and Y.H. Yan. 2018. 'Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns', *Gigascience*, 7: 1-11.

Shih, P.M., and N.J. Matzke. 2013. 'Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins', *Proceedings of the National Academy of Sciences USA*, 110: 12355-60.

Sidow, A. 1996. 'Gen(om)e duplications in the evolution of early vertebrates', *Current Opinion in Genetics & Development*, 6: 715-22.

Silvestro, D., B. Cascales-Minana, C.D. Bacon, and A. Antonelli. 2015. 'Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record', *New Phytologist*, 207: 425-36.

Simonin, K.A., and A.B. Roddy. 2018. 'Genome downsizing, physiological novelty, and the global dominance of flowering plants', *PLOS Biology*, 16: e2003706.

Smith, S.A., J.W. Brown, Y. Yang, R. Bruenn, C.P. Drummond, S.F. Brockington, J.F. Walker, N. Last, N.A. Douglas, and M.J. Moore. 2018. 'Disparity, diversity, and duplications in the Caryophyllales', *New Phytologist*, 217: 836-54.

Smith, S.Y., R.A. Stockey, G.W. Rothwell, and S.A. Little. 2017. 'A new species of *Pityostrobus* (Pinaceae) from the Cretaceous of California: moving towards understanding the Cretaceous radiation of Pinaceae', *Journal of Systematic Palaeontology*, 15: 69-81.

Soltis, D.E., C.J. Visger, and P.S. Soltis. 2014a. 'The polyploidy revolution then...and now: Stebbins revisited', *American Journal of Botany*, 101: 1057-78.

Soltis, D.E., V.A. Albert, J. Leebens-Mack, C.D. Bell, A.H. Paterson, C. Zheng, D. Sankoff, C.W. Depamphilis, P.K. Wall, and P.S. Soltis. 2009. 'Polyploidy and angiosperm diversification', *American Journal of Botany*, 96: 336-48.

Soltis, D.E., M.C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N.M. Miles, E.V. Mavrodiev, W. Mei, M.B. Cortez, P.S. Soltis, and M.A. Gitzendanner. 2014b.

'Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. ()', *New Phytologist*, 202: 1105-17.

Soltis, P.S., and D.E. Soltis. 2016. 'Ancient WGD events as drivers of key innovations in angiosperms', *Current Opinion in Plant Biology*, 30: 159-65.

Soreng, R.J., and J.I. Davis. 1998. 'Phylogenetics and character evolution in the grass family (Poaceae): Simultaneous analysis of morphological and chloroplast DNA restriction site character sets', *Botanical Review*, 64: 1-85.

Spoelhof, J.P., P.S. Soltis, and D.E. Soltis. 2017. 'Pure polyploidy: Closing the gaps in autopolyploid research', *Journal of Systematics and Evolution*, 55: 340-52.

Stanich, N.A., G.W. Rothwell, and R.A. Stockey. 2009. 'Phylogenetic diversification of *Equisetum* (Equisetales) as inferred from Lower Cretaceous species of British Columbia, Canada', *American Journal of Botany*, 96: 1289-99.

Stebbins, G.L. 1947. 'Types of Polyploids: Their Classification and Significance.' in M. Demerec (ed.), *Advances in Genetics* (Academic Press).

Stebbins Jr, G.L. 1940. 'The significance of polyploidy in plant evolution', *The American Naturalist*, 74: 54-66.

Steemans, P., A.L. Herisse, J. Melvin, M.A. Miller, F. Paris, J. Verniers, and C.H. Wellman. 2009. 'Origin and radiation of the earliest vascular land plants', *Science*, 324: 353.

Steige, K.A., and T. Slotte. 2016. 'Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy', *Current Opinion in Plant Biology*, 30: 88-93.

Stewart, W.N., and G.W. Rothwell. 1993. *Paleobotany and the evolution of plants* (Cambridge University Press).

Stolzer, M., H. Lai, M. Xu, D. Sathaye, B. Vernot, and D. Durand. 2012. 'Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees', *Bioinformatics*, 28: i409-i15.

Szollosi, G.J., E. Tannier, N. Lartillot, and V. Daubin. 2013. 'Lateral gene transfer from the dead', *Systematic Biology*, 62: 386-97.

Szollosi, G.J., B. Boussau, S.S. Abby, E. Tannier, and V. Daubin. 2012. 'Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations', *Proceedings of the National Academy of Sciences USA*, 109: 17513-8.

Szöllősi, G.J., A.A. Davín, E. Tannier, V. Daubin, and B. Boussau. 2015. 'Genome-scale phylogenetic analysis finds extensive gene transfer among fungi', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370.

Tang, H., J.E. Bowers, X. Wang, and A.H. Paterson. 2010. 'Angiosperm genome comparisons reveal early polyploidy in the monocot lineage', *PNAS*, 107.

Tang, H., X. Wang, J.E. Bowers, R. Ming, M. Alam, and A.H. Paterson. 2008. 'Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps', *Genome Research*, 18: 1944-54.

Tank, D.C., J.M. Eastman, M.W. Pennell, P.S. Soltis, D.E. Soltis, C.E. Hinchliff, J.W. Brown, E.B. Sessa, and L.J. Harmon. 2015. 'Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications', *New Phytologist*, 207: 454-67.

Taylor, E.L., T.N. Taylor, and M. Krings. 2009. *Paleobotany: the biology and evolution of fossil plants* (Academic Press).

Taylor, J.S., and J. Raes. 2004. 'Duplication and divergence: the evolution of new genes and old ideas', *Annual Review of Genetics*, 38: 615-43.

te Beest, M., J.J. Le Roux, D.M. Richardson, A.K. Brysting, J. Suda, M. Kubesova, and P. Pysek. 2012. 'The more the better? The role of polyploidy in facilitating plant invasions', *Annals of Botany*, 109: 19-45.

Teufel, A.I., L. Liu, and D.A. Liberles. 2016. 'Models for gene duplication when dosage balance works as a transition state to subsequent neo-or sub-functionalization', *BMC Evolutionary Biology*, 16: 45.

Theissen, G., R. Melzer, and F. Rumpler. 2016. 'MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution', *Development*, 143: 3259-71.

Thomas, G.H., and R.P. Freckleton. 2012. 'MOTMOT: models of trait macroevolution on trees', *Methods in Ecology and Evolution*, 3: 145-51.

Thompson, A., H.H. Zakon, and M. Kirkpatrick. 2016. 'Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes', *Genetics*, 202: 765-74.

Thorne, J.L., and H. Kishino. 2005. 'Estimation of divergence times from molecular sequence data.' in Rasmus Nielsen (ed.), *Statistical methods in molecular evolution* (Springer: New York).

Tiley, G.P., C. Ane, and J.G. Burleigh. 2016. 'Evaluating and characterizing ancient whole-genome duplications in plants with gene count data', *Genome Biology and Evolution*, 8: 1023-37.

Tomescu, A.M., S.E. Wyatt, M. Hasebe, and G.W. Rothwell. 2014. 'Early evolution of the vascular plant body plan - the missing mechanisms', *Current Opinion in Plant Biology*, 17: 126-36.

Tomescu, A.M., I.H. Escapa, G.W. Rothwell, A. Elgorriaga, and N.R. Cuneo. 2017. 'Developmental programmes in the evolution of Equisetum reproductive morphology: a hierarchical modularity hypothesis', *Annals of Botany*, 119: 489-505.

Twyford, A.D. 2018. 'The road to 10,000 plant genomes', *Nature Plants*, 4: 312-13.
Valentine, J.W., B.H. Tiffney, and J.J. Sepkoski, Jr. 1991. 'Evolutionary dynamics of plants and animals: a comparative approach', *Palaios*, 6: 81-8.
Van Bel, M., T. Diels, E. Vancaester, L. Kreft, A. Botzki, Y. Van de Peer, F. Coppens, and K. Vandepoele. 2018. 'PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics', *Nucleic Acids Res*, 46: D1190-D96.
Van de Peer, Y., S. Maere, and A. Meyer. 2010. '2R or not 2R is not the question anymore', *Nature Reviews Genetics*, 11: 166.
Van de Peer, Y., E. Mizrachi, and K. Marchal. 2017. 'The evolutionary significance of polyploidy', *Nature Reviews Genetics*, 18: 411-24.
Vandenberghe, N., F.J. Hilgen, and R.P. Speijer. 2012. 'The Paleogene Period.' in F. M. Gradstein, J. G. Ogg, M. Schmitz and G. Ogg (eds.), *The geologic timescale 2012* (Elsevier: Ansterdam).
Vandenbroucke, T.R.A., M. Williams, J.A. Zalasiewicz, J.R. Davies, and R.A. Waters. 2008. 'Integrated Upper Ordovician graptolite-chitinozoan biostratigraphy of the Cardigan and Whitland areas, southwest Wales', *Geological Magazine*, 145: 199-214.
Vanneste, K., Y. Van de Peer, and S. Maere. 2013. 'Inference of genome duplications from age distributions revisited', *Molecular Biology and Evolution*, 30: 177-90.
Vanneste, K., S. Maere, and Y. Van de Peer. 2014a. 'Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369.
Vanneste, K., G. Baele, S. Maere, and Y. Van de Peer. 2014b. 'Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary', *Genome Research*, 24: 1334-47.
Vanneste, K., L. Sterck, A.A. Myburg, Y. Van de Peer, and E. Mizrachi. 2015. 'Horsetails are ancient polyploids: evidence from *Equisetum giganteum*', *Plant Cell*, 27: 1567-78.
Veitia, R.A. 2004. 'Gene dosage balance in cellular pathways: implications for dominance and gene duplicability', *Genetics*, 168: 569-74.
Veitia, R.A., S. Bottani, and J.A. Birchler. 2008. 'Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects', *Trends in Genetics*, 24: 390-7.
Veron, A.S., K. Kaufmann, and E. Bornberg-Bauer. 2007. 'Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins', *Molecular Biology and Evolution*, 24: 670-8.

Villarreal, J.C., D.C. Cargill, A. Hagborg, L. Soderstrom, and K.S. Renzaglia. 2014. 'A synthesis of hornwort diversity: Patterns, causes and future work', *Phytotaxa*, 9: 17.

Vivancos, J., L. Spinner, C. Mazubert, F. Charlot, N. Paquet, V. Thareau, M. Dron, F. Nogue, and C. Charon. 2012. 'The function of the RNA-binding protein TEL1 in moss reveals ancient regulatory mechanisms of shoot development', *Plant Molecular Biology*, 78: 323-36.

Wagner, G.P., C. Amemiya, and F. Ruddle. 2003. 'Hox cluster duplications and the opportunity for evolutionary novelties', *Proceedings of the National Academy of Sciences USA*, 100: 14603-06.

Wagner, P.J. 2010. 'Paleontological perspectives on morphological evolution.' in M.A Bell, Futuyma, D.J.., Eanes, W.F, Levinton, J.S (ed.), *Evolution since Darwin: the First 150 Years* (Sinauer Associates).

Walker, J.F., Y. Yang, M.J. Moore, J. Mikenas, A. Timoneda, S.F. Brockington, and S.A. Smith. 2017. 'Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales', *American Journal of Botany*, 104: 858-67.

Wan, T., Z.M. Liu, L.F. Li, A.R. Leitch, I.J. Leitch, R. Lohaus, Z.J. Liu, H.P. Xin, Y.B. Gong, Y. Liu, W.C. Wang, L.Y. Chen, Y. Yang, L.J. Kelly, J. Yang, J.L. Huang, Z. Li, P. Liu, L. Zhang, H.M. Liu, H. Wang, S.H. Deng, M. Liu, J. Li, L. Ma, Y. Liu, Y. Lei, W. Xu, L.Q. Wu, F. Liu, Q. Ma, X.R. Yu, Z. Jiang, G.Q. Zhang, S.H. Li, R.Q. Li, S.Z. Zhang, Q.F. Wang, Y. Van de Peer, J.B. Zhang, and X.M. Wang. 2018. 'A genome for gnetophytes and early evolution of seed plants', *Nature Plants*, 4: 82-89.

Wang, Y., S.P. Ficklin, X. Wang, F.A. Feltus, and A.H. Paterson. 2016. 'Large-scale gene relocations following an ancient genome triplication associated with the diversification of core eudicots', *PLoS One*, 11: e0155637.

Warnock, R.C.M., Z. Yang, and P.C.J. Donoghue. 2012. 'Exploring uncertainty in the calibration of the molecular clock', *Biology Letters*, 8: 156-59.

Warnock, R.C.M., J.F. Parham, W.G. Joyce, T.R. Lyson, and P.C.J. Donoghue. 2015. 'Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors', *Proceedings of the Royal Society B: Biological Sciences*, 282: 20141013.

Wickett, N.J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, M.S. Barker, J.G. Burleigh, M.A. Gitzendanner, B.R. Ruhfel, E. Wafula, J.P. Der, S.W. Graham, S. Mathews, M. Melkonian, D.E. Soltis, P.S. Soltis, N.W. Miles, C.J. Rothfels, L. Pokorny, A.J. Shaw, L. DeGironimo, D.W. Stevenson, B. Surek, J.C. Villarreal, B. Roure, H. Philippe, C.W. dePamphilis,

T. Chen, M.K. Deyholos, R.S. Baucom, T.M. Kutchan, M.M. Augustin, J. Wang, Y. Zhang, Z. Tian, Z. Yan, X. Wu, X. Sun, G.K. Wong, and J. Leebens-Mack. 2014. 'Phylotranscriptomic analysis of the origin and early diversification of land plants', *Proc Natl Acad Sci U S A*, 111: E4859-68.

Wilf, P., and K.R. Johnson. 2004. 'Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafloral record', *Paleobiology*, 30: 347-68.

Wilhelmsson, P.K.I., C. Muhlich, K.K. Ullrich, and S.A. Rensing. 2017. 'Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in streptophyte algae', *Genome Biology and Evolution*, 9: 3384-97.

Wills, M.A. 2001. 'Morphological disparity: A primer.' in Jonathan M. Adrain, Gregory D. Edgecombe and Bruce S. Lieberman (eds.), *Fossils, Phylogeny, and Form* (Springer US: Boston, MA).

Wills, M.A., D.E.G. Briggs, and R.A. Fortey. 1994. 'Disparity as an evolutionary index - a comparison of Cambrian and recent arthropods', *Paleobiology*, 20: 93-130.

Wood, T.E., N. Takebayashi, M.S. Barker, I. Mayrose, P.B. Greenspoon, and L.H. Rieseberg. 2009. 'The frequency of polyploid speciation in vascular plants', *Proceedings of the National Academy of Sciences USA*, 106: 13875-9.

Xiang, Y., C.H. Huang, Y. Hu, J. Wen, S. Li, T. Yi, H. Chen, J. Xiang, and H. Ma. 2017. 'Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication', *Molecular Biology and Evolution*, 34: 262-81.

Xiong, Z., R.T. Gaeta, and J.C. Pires. 2011. 'Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid Brassica napus', *Proceedings of the National Academy of Sciences USA* 108: 7908-13.

Yang, Z. 2007. 'PAML 4: phylogenetic analysis by maximum likelihood', *Molecular Biology and Evolution*, 24: 1586-91.

Yang, Z., and B. Rannala. 2006. 'Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds', *Molecular Biology and Evolution*, 23: 212-26.

Yu, Y., Q. Xiang, P.S. Manos, D.E. Soltis, P.S. Soltis, B.H. Song, S. Cheng, X. Liu, and G. Wong. 2017. 'Whole-genome duplication and molecular evolution in Cornus L. (Cornaceae) - Insights from transcriptome sequences', *PLoS One*, 12: e0171361.

Zanne, A.E., D.C. Tank, W.K. Cornwell, J.M. Eastman, S.A. Smith, R.G. FitzJohn, D.J. McGlinn, B.C. O'Meara, A.T. Moles, P.B. Reich, D.L. Royer, D.E. Soltis,

P.F. Stevens, M. Westoby, I.J. Wright, L. Aarssen, R.I. Bertin, A. Calaminus, R. Govaerts, F. Hemmings, M.R. Leishman, J. Oleksyn, P.S. Soltis, N.G. Swenson, L. Warman, and J.M. Beaulieu. 2014. 'Three keys to the radiation of angiosperms into freezing environments', *Nature*, 506: 89-92.

Zwieniecki, M.A., and C.K. Boyce. 2014. 'Evolution of a unique anatomical precision in angiosperm leaf venation lifts constraints on vascular plant ecology', *Proceedings of the Royal Society B: Biological Sciences*, 281: 20132829.

Appendix 1. Taxa included in the molecular clock analysis to date the *zeta* and *epsilon* WGD event and their source.

Species	Order	Source	Species	Order	Source
Sphagnum lescurii	Sphagnales	1KP	*Ophioglossum petiolatum*	Ophioglossales	1KP
Physcomitrella patens	Funariales	Plaza 4.0	*Angiopteris evecta*	Marattiales	1KP
Ceratodon purpureas	Dicranales	1KP	*Alsophila spinulosa*	Cyatheales	1KP
Hedwigia ciliata	Hedwigiales	1KP	*Pteridium aquilinum*	Polypodiales	1KP
Thuidium delicatulum	Hypnales	1KP	*Picea abies*	Pinales	GreenPhyl 4.0
Leucodon sciuroides	Hypnales	1KP	*Pinus taeda*	Pinales	1KP
Anomodon attenuates	Hypnales	1KP	*Cedrus libani*	Pinales	1KP
Rhynchostegium serrulatum	Hypnales	1KP	*Prumnopitus andina*	Pinales	1KP
Bryum argenteum	Bryales	1KP	*Cunninghamia lanceolata*	Pinales	1KP
Rosulabryum capillare	Bryales	1KP	*Juniperus scopulorum*	Pinales	1KP
			Taxus baccata	Pinales	1KP
Nothoceros aegnimaticus	Dendroceratales	1KP	*Sciadopitys verticillata*	Pinales	1KP
Nothoceros vincentianus	Dendroceratales	1KP	*Zamia vasquezii*	Cycadales	1KP
			Cycas mycolitzii	Cycadales	1KP
Marchantia polymorpha	Marchantiales	1KP	*Ginkgo biloba*	Ginkgoales	1KP
Marchantia emarginata	Marchantiales	1KP	*Ephedra sinica*	Gnetales	1KP
Ricciocarpos natans	Marchantiales	1KP	*Gnetum montanum*	Gnetales	1KP
Sphaerocarpos texanus	Sphaerocarpales	1KP	*Welwitschia mirabilis*	Gnetales	1KP
Bazzania trilobata	Jungermanniales	1KP			
Metzgeria crassipilis	Metzgeriales	1KP	*Amborella trichopoda*	Amborellales	Plaza 4.0
			Nuphar advena	Nymphaeales	1KP
Selaginella moellendorfii	Selaginellales	Plaza 4.0	*Kadsura heteroclite*	Austrobaileyales	1KP
Selaginella stauntoniana	Selaginellales	1KP	*Houttuynia cordata*	Piperales	1KP
Huperzia squarrosa	Lycopodiales	1KP	*Saruma henryi*	Piperales	1KP
Pseudolycopodiella caroliana	Lycopodiales	1KP	*Liriodendron tulipfera*	Magnoliales	1KP
Dendrolycopodium obscurum	Lycopodiales	1KP	*Persea americana*	Laurales	1KP
			Sarcandra glabra	Chloranthales	1KP
Equisetum diffusum	Equisetales	1KP	*Acorus americanus*	Acorales	1KP
Psilotum nudum	Psilotales	1KP			

Dioscorea villosa	Dioscoreales	1KP
Smilax bona-nox	Liliales	1KP
Colchicum autumnale	Liliales	1KP
Yucca filamentosa	Asparagales	1KP
Sabal bermudana	Arecales	1KP
Elaeis guineensis	Arecales	GreenPhyl 4.0
Phoenix dactylifera	Arecales	GreenPhyl 4.0
Musa acuminata	Zingiberales	Plaza 4.0
Musa balbisiana	Zingiberales	Phytozome
Oryza sativa	Poales	Plaza 4.0
Panicum hallii	Poales	Phytozome
Hordeum vulgare	Poales	Plaza 4.0
Sorghum bicolor	Poales	Plaza 4.0
Seteria italica	Poales	Plaza 4.0
Zea mays	Poales	Plaza 4.0
Brachypodium distachyon	Poales	Plaza 4.0
Escholzia californicum	Ranunculales	1KP
Aquiliegia formosa	Ranunculales	1KP
Podophyllum peltatum	Ranunculales	1KP
Beta vulgaris	Caryophyllales	Plaza 4.0
Diospyros malabarica	Ericales	1KP
Inula helenium	Asterales	1KP
Tanacetum parthenium	Asterales	1KP
Ipomoea purpurea	Solanales	1KP
Solanum tuberosum	Solanales	Plaza 4.0
Solanum lycospersicum	Solanales	Plaza 4.0
Rosmarinus officinales	Lamiales	1KP
Mimulus guttatus	Lamiales	Phytozome
Catharanthus roseus	Gentianales	1KP
Coffea canephora	Gentianales	GreenPhyl 4.0
Allamanda cathartica	Gentianales	1KP
Vitis vinifera	Vitales	Plaza 4.0
Ricinus communis	Malpighiales	Plaza 4.0
Manihot esculenta	Malpighiales	Plaza 4.0
Cucumis melo	Cucurbitales	Plaza 4.0
Cucumis sativus	Cucurbitales	Plaza 4.0
Citrullus lanatus	Cucurbitales	Plaza 4.0
Larrea tridentata	Rosales	1KP
Fragaria vesca	Rosales	Plaza 4.0
Prunus persica	Rosales	Plaza 4.0
Malus domestica	Rosales	Plaza 4.0
Boehmeria nivea	Rosales	1KP
Lotus japonicus	Fabales	Plaza 4.0
Cicer arietinum	Fabales	GreenPhyl 4.0
Cajanus cajan	Fabales	GreenPhyl 4.0
Glycine max	Fabales	Plaza 4.0
Medicago truncatula	Fabales	Plaza 4.0
Eucalyptus grandis	Myrtales	Plaza 4.0
Citrus sinensis	Sapindales	Plaza 4.0
Gossypium raimondii	Malvales	Plaza 4.0
Hibiscus cannabinus	Malvales	1KP
Theobroma cacao	Malvales	Plaza 4.0
Arabidopsis thaliana	Brassicales	Plaza 4.0
Arabidopsis lyrata	Brassicales	Plaza 4.0
Capsella rubella	Brassicales	Plaza 4.0
Capsella grandiflora	Brassicales	Phytozome
Brassica rapa	Brassicales	Plaza 4.0
Thelungiella parvula	Brassicales	Plaza 4.0
Eutrema salsugineum	Brassicales	Phytozome
Boechera stricta	Brassicales	Phytozome
Linum usitatissimum	Malpighiales	Phytozome
Populus trichocarpa	Malpighiales	Plaza 4.0

Appendix 2. A full list of fossil calibrations updated for use in molecular clock analysis. All seed plants calibrations were applied twice across the tree, and each angiosperm calibration applied four times. Ages in millions of years before present. Most of these calibrations were devised by Jose Barba-Montoya (Barba-Montoya *et al.* 2018) and John Clarke (Clarke *et al.* 2011) and much of the text used has been taken from their work. Where necessary, they were all revised by the author and updated, but as such they are not presented as part of Chapter 2.

1. CG Embryophytes | MRCA: *Marchantia* – *Capsella* | 448.5 – 509 Ma
Fossil taxon and specimen. Following Clarke *et al.*[1], constraints were based on trilete spores from the Qusaiba-1 core from the Quasim formation of northern Saudi Arabia[2] and Cambrian spores of the Bright Angel Shale in the lower elevations of the Grand Canyon, Arizona[3]
Phylogenetic justification. Following Clarke *et al.*[1], the oldest records of liverworts date to the Early Devonian, however trilete spores support the total group Anthocerotae + Tracheophyta, providing a minimum constraint. The Cambrian spores of the Bright Angel Shale represent the oldest spores possessing two Embryophyte synapomorphies: permanent dyad and tetrad arrangements and multilamellate sporoderm.
Minimum age. 448.5 Ma.
Maximum age. 509 Ma.
Age justification. The minimum constraint, following Clarke et al.[1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia. We follow Clarke *et al.*[1] and accept a likely minimum age at the top of the *Acanthochitina barbata* biozone based on co-occurrence[3], the base of which is estimated at 448.5 Ma., following Cooper *et al.*[4]. The maximum constraint is based on the Cambrian spores of the Bright Angel Shale, which falls fully within the span of the *Albertella, Glossopluera* and *Ehmaniella* trilobite biozones, representing 507.2-509 Ma.[5]

2. CG Marchantiopsida | MRCA *Sphaerocarpos* – *Marchantia* | 228.4 Ma
Fossil taxon and specimen. *Marchantites cyatheoides* [Plate 1A. number 13929 South African Museum Cape Town] from the Upper Umkomaas, Natal, Molteno Formation
Phylogenetic justification. Orginially assigned to the broad genus *Hepaticites* by Townrow[6], however Anderson[7] revised the taxon and placed it within the genus *Marchantites* based on the presence of a prostate forked thallus, a conspicuous midrib, rhizoides, air chambers and central scales, all indicating an affinity with the Marchantiaceae
Minimum age. 228.4 Ma.
Age justification. *Marchantites cyatheoides* is known only from the Molteno formation of South Africa and the Middle Triassic Sydney basin, Australia. The Molteno formation is among the most intensely studied Upper Triassic formations in the world, and based on the megaflora assemblages, was dated as Carnian by Anderson & Anderson[8]. As no formal boundary is defined for the Molteno formation, we took the upper boundary of the Carnian following Ogg[9] as 228.4 Ma.

3. SG Metzgeriales | MRCA *Bazzania* – *Metzgeria* | 407.6 Ma.]
Fossil taxon and specimen. *Riccardiothallus devonicus* [CBY_n9004008 Museum of Plant History, Institute of Botany, Chinese Academy of Sciences] from the Posongchong formation, Zhichang Village, Gumu Town, Wenshan District, Yunnan Province, China.

Phylogenetic justification. Guo et al.[10] determined that *Riccardiothallus* shares several similarities with the extant genus *Riccardia* (Aneuraceae), including a flattened thallus with irregular branching, lack of conducting tissue and a lack of a costa, yet based on the age of the fossil, it was deemed most appropriate to assign it to a new genus.
Minimum age. 407.6 Ma.
Age justification. *Riccardiothallus* comes from the Posongchong formation in China, the stratigraphy of which was confirmed by Hao et al.[11] as Lower Devonian (Pragian), based on the evidence of marine invertebrates from the overlying Pojiao formation. The upper limit of the Pragian (407.6 Ma.) was adopted as the minimum age following Becker *et al.*[12]

4. CG Stomatophyta | MRCA: *Sphagnum* – Tracheophyta + Anthocerophyta | 426.7 – 509 Ma
Fossil taxon and specimen. Following Clarke *et al.*[1] *Cooksonia cambrensis* [TCD22951, Department of Geology, Trinity College, Dublin] from the Devilsbit Mountain Area, Central Ireland was accepted as the oldest representative of total group Tracheophyta
Phylogenetic justification. The fossil record of mosses is poor and *Sporogonites* remains the oldest possible moss, though its phylogenetic position is too equivocal to provide a minimum constraint and so following Clarke *et al.*[1] *Cooksonia* was used to provide a minimum constraint, having been reinterpreted as a member of total group Tracheophyta rather than crown group[1], on the basis that many of the characters that placed *Cooksonia* in the crown group are found only in younger specimens, and some of the characters, such as the presence of the sterome, are unlikely to be synapomorphies of the crown group[13]. Placing *Cooksonia* in the total group is congruent with unequivocal total group synapomorphies, such as multiple sporangia and differentially thickened tracheids[13].
Minimum age. 426.7 Ma.
Maximum age. 509 Ma.
Age justification. Following Clarke *et al.*[1] the earliest occurrences of *Cooksonia* are bracketed by graptolites that are characteristic of the *ludensis* biozone, which coincides with the Wenlock-Ludlow series boundary[14], providing a minimum age of 426.7 Ma. updated following Melchin *et al.*[15]. Also following Clarke *et al.*[1] the oldest members of total group Tracheophyta would likely have shared the poor fossilization characteristics as Bryophyte grade material, and is likely a poor approximation of the age of the clade, and so we followed a soft maximum age of 509 Ma.

5. SG Bryidae | MRCA: *Thuidium* – *Bryum* | 259.7 Ma.
Fossil taxon and specimen. *Campimirinus riopratense* [UNICAMP: CP1/155-195 at the University of Campinas] Teresina Formation (Permian–Guadalupian) collected in the Rio Preto Quarry in the state of Paraná, southern Brazil.
Phylogenetic justification. Though likened to the modern genus *Hypnum,* De Souza et al.[16] were reluctant to assign *C. riopratense* to an extant clade based on the absence of double short costae in the gametophyte and other key diagnostic features and so favoured the creation of a new genus. Following Laenen et al.[17] it was assigned to the Hypnales based on the similarity to early divergent pleurocarpous mosses.
Minimum age. 259.8
Age justification. Following De Souza *et al.*[16], the Teresina Formation falls within the Passa Dois Group. Based on U/Pb isotopes, Santos *et al.*[18] established the base age of this group as 270.6 +/- 0.7

Ma. As no formal upper boundary for the Terasina formation is established, a minimum age was constructed based on the upper boundary of the Guadalupian at 260.4 +/- 0.7 following Davydov *et al.*[19]

6. MRCA: *Nothoceros* – *Huperzia* | 426.7 – 509 Ma.
Fossil taxon and specimen. Following Clarke *et al.*[1], *Cooksonia cambrensis* [TCD22951, Department of Geology, Trinity College, Dublin] from the Devilsbit Mountain Area, Central Ireland was accepted as the oldest representative of total group Tracheophyta
Phylogenetic justification. Following Clarke *et al.*[1], *Cooksonia* was reinterpreted as a member of total group Tracheophyta rather than crown group, on the basis that many of the characteris that placed *Cooksonia* in the crown group are found only in younger specimens, and some of the characters, such as the presence of the sterome, are unlikely to be synapomorphies of the crown group[13]. Placing *Cooksonia* in the total group is congruent with unequivocal total group synapomorphies, such as multiple sporangia and differentially thickened tracheids[13].
Minimum age. 426.7 Ma.
Soft maximum age. 509 Ma.
Age justification. Following Clarke *et al.*[1] the earliest occurrences of *Cooksonia* are bracketed by graptolites that are characteristic of the *ludensis* biozone, which coincides with the Wenlock-Ludlow series boundary[14], providing a minimum age of 426.7 Ma. updated following Melchin *et al.*[15]. Also following Clarke *et al.*[1] the oldest members of total group Tracheophyta would likely have shared the poor fossilization characteristics as Bryophyte grade material, and is likely a poor approximation of the age of the clade, and so we followed a soft maximum age of 509 Ma.

7. CG Tracheophyta | MRCA: Lycophyta-Euphyllophyta | 422 Ma – 449.6 Ma.
Fossil taxon and specimen. Clarke *et al.*[1] based their calibration of this node on *Zosterophyllum* sp. [US384-8137; University of Saskatchewan Collections, Canada] from Bathurst Island[20].
Phylogenetic justification. Following Clarke *et al.*[1] the *Zosterophyllum* sp. from Bathurst Island (Kotyk *et al.*[20]) is unequivocally zostrophyll given its possession of reniform sporangia, sporangia that dehisce along their distal margins, and laterally inserted sporangia. All *Zosterophyllum* species are total group Lycopsida[13].
Minimum age. 422 Ma.
Soft maximum age. 449.5 Ma.
Age justification. *Zosterophyllum* sp. on Bathurst Island[20] co-occurs with conodont *Ozarkodina douroensis*, which is restricted to the Ludlow (as O. n. sp. B in[4-7]. Thus, a minimum age interpretation can be derived from the top of the Ludlow, dated to 423.0 Ma ± 1.0 Myr, thus 422.0 Ma. The soft maximum constraint, following Clarke et al.[1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia. We follow Clarke *et al.*[1] and accept a likely a softa maximum at the top of the *Acanthochitina barbata* biozone based on co-occurrence[3], the base of which is estimated at 449.5 Ma, following Cooper *et al.*[4].

8. CG Lycophytes | MRCA: *Huperzia-Selaginella* | 392.1 Ma – 449.5 Ma.
Fossil taxon and specimen. *Leclercquia complexa* [CW092 (07 – 061): Collections of the Centre for Palynological Studies, Department of Animal and Plant Sciences, University of Sheffield, UK], from Campbellton Formation outcropping on the south shore of the Restigouche River, between Dalhousie and Campbellton, New Brunswick, eastern Canada[21].

Phylogenetic justification. Kenrick and Crane[3] identified *Leclercquia complexa* as the oldest member of Isoetopsida and crown Lycopodiophyta. This interpretation is supported by spore characteristics analysed phylogenetically by Wellman et al.[21].
Minimum age. 392.1 Ma.
Soft Maximum age. 449.5 Ma.
Age justification. A Late Emsian age is often cited for the New Brunswick occurrences of identified *Leclercquia complexa* e.g. [22] and, indeed, the *Stockmensella-Leclerqia* macroplant Biozone spans all but the earliest Emsian[12]. However, Wellman et al.[22] attribute their own material of *Leclercquia complexa* to the middle of the *Emphanisporites annulatus – Camarozonotriletes sextantii* Spore Assemblage Biozone which falls within the early part of the Emsian. In either instance, the earliest records of *Leclercquia complexa* fall fully within the Emsian, the end of which is dated to 393.3 Ma ± 1.2 Myr[12], yielding a minimum constraint of 392.1 Ma. The soft maximum constraint, following Clarke et al.[1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia. We follow Clarke *et al.*[1] and accept a likely a softa maximum at the top of the *Acanthochitina barbata* biozone based on co-occurrence[3], the base of which is estimated at 449.5 Ma., following Cooper *et al.*[4].
Discussion. Magallon et al.[23] cite a minimum age of 385 Ma, based on the Middle-Upper Devonian Boundary, but our more detailed stratigraphy allows for an older minimum age interpretation of *Leclercquia complexa.*

9. CG Euphyllophytes | MRCA: Monilophyta-Spermatophyta | 385.571 Ma – 449.5 Ma.
Fossil taxon and specimen. *Rellimia thomsonii* from the Panther Mountain Formation of New York[24] [335.34; Paleobotanical Collection of the State University of New York at Bingham].
Phylogenetic justification. Magallón et al.[23] identified *Ibyka amphikoma*[25] as the oldest record of the pteridophyte lineage based on phylogenetic analyses undertaken by Kenrick and Crane[13].
Minimum age. 384.71 Ma.
Soft maximum age. 449.5 Ma.
Age justification. Clarke et al.[1] proposed *Rellimia thomsonii*, an aneurophytalean progymnosperm from the Panther Mountain Formation of New York[24], as the oldest record of crown Euphyllophyta. The Panther Mountain Formation is equivalent to the Ludlowville and Skaneateles formations[1], which occur below the Moscow Formation of New York[26], making *Rellimia thomsonii* older than *Ibyka amphikoma*[1]. The Ludlowville-Moscow formation boundary falls deep within the Lower *varcus* zone[27] and, therefore, below the *rhenanus-ansatus* biozonal boundary[12], at the very least, which has been dated to 386.25 Ma ± 0.679 Myr, yielding a minimum constraint of 385.571 Ma. The soft maximum constraint, following Clarke et al.[1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia. We follow Clarke *et al.*[1] and accept a likely a softa maximum at the top of the *Acanthochitina barbata* biozone based on co-occurrence[3], the base of which is estimated at 449.5 Ma., following Cooper *et al.*[4].
Discussion. Magallón et al.[23] established a minimum age constraint using *Ibyka amphikoma,* based on the Givetian-Frasnian boundary, for which they provided a date of 385 Ma, though this has since been revised to 382.7 Ma ± 1 Myr[12]. *Ibyka amphikoma* was recovered from the Manorkill Shale Member, which is a lateral equivalent of the Windom Member, within the Moscow Formation of New York[28,29], which falls fully within the *ansatus* conodont Biozone[30,31] the top of which is dated to 385.41 Ma ± 0.7 Myr[12], thus, yielding a minimum age constraint of 384.71 Ma, younger than the minimum age of *Rellimia thomsonii*.

10. CG Monilophytes | MRCA: *Equisetum* - *Pteridium* | 384.71 Ma – 449.5 Ma.
Fossil taxon and specimen. *Ibyka amphikoma* was recovered from the Manorkill Shale Member at Schoharie Creek directly below the spillway of Gilboa dam, Gilboa, Schoharie County, New York, Gilboa[25].
Phylogenetic justification. *Ibyka amphikoma*[25] is the oldest record of the equisetopsid lineage based on the phylogenetic analyses undertaken by Kenrick and Crane[13].
Minimum age. 384.71 Ma.
Soft Maximum age. 449.5 Ma.
Age justification. *Ibyka amphikoma* was recovered from the Manorkill Shale Member, which is a lateral equivalent of the Windom Member, within the Moscow Formation of New York[28,29], which falls fully within the *ansatus* conodont Biozone[30,31] the top of which is dated to 385.41 Ma ± 0.7 Myr [12], thus, yielding a minimum age constraint of 384.71 Ma. The soft maximum constraint, following Clarke et al.[1], is based on the oldest occurrences of trilete spores, known from the Qusaiba-1 core from the Quasim Formation of northern Saudi Arabia. We follow Clarke *et al.*[1] and accept a likely a softa maximum at the top of the *Acanthochitina barbata* biozone based on co-occurrence[3], the base of which is estimated at 449.5 Ma., following Cooper *et al.*[4].
Discussion. Magallón et al.[23] established a minimum age constraint based on *Ibyka amphikoma* using the Givetian-Frasnian boundary, for which they provided a date of 385 Ma, though this has since been revised to 382.7 Ma ± 1 Myr[12]. However, we provide a more detailed stratigraphic justification for the age of *I. amphikoma* which allows for an older minimum age constraint.

11. SG Leptosporangiate ferns | MRCA: *Angiopteris* – *Pteridium* | 315.1 Ma.
Fossil taxon and specimen. *Senftenbergia plumosa* [E3672, National Museum, Prague] from the Kladno formation of the Nyrany locality in the Pilsen Basin, Bohemian Massif[32]
Phylogenetic justification. Despite similar reproductive tissues to members of the Schizeaceae, *Senftenbergia plumosa* assigned to the Tedeleaceae based on angular diametric cells following Pšenička and Bek[32] following careful examination of the epidermal cells and cuticular layer.
Minimum age. 315.1 Ma.
Age justification. *S. plumosa* occurs throughout the Westphalian A to the Lower Permian following Bek and Pšenička[33], and so the upper limit of the Westphalian A was accepted as a minimum constraint. Unfortunately, the boundary of the Westphalian A does not correlate with the current Geologic Time Scale, and so the upper boundary of the Westphalian B (315.2 +/- 0.1) was taken as the minimum age following Davydov et al.[19]

12. SG Polypodiales | MRCA: *Alsophila* – *Pteridium* | 98.79 Ma.
Fossil taxon and specimen. *Kramerop teris resinatus* [AMNH Bu-ASJH-3] from Amber mines near Tanai in Kachin State, Myanmar[34].
Phylogenetic justification. Schmidt et al.[34] assigned *K. resinatus* to the Dennstaedtiaceae based on the presence of polypod sporangia, free-veined leaves and exindusiate sori. However, irregular tuber shaped structures on the leaves are unique among extant ferns and so it was assigned to its own genus[34].
Minimum age. 98.79 Ma.

Age justification. Biostratigraphic studies suggested a late Albian age of the amber-bearing sediment (Cruickshank and Ko[35]) hence the inclusions have a late Early Cretaceous age, with a minimum age of 98.79 million years (earliest Cenomanian, early Late Cretaceous) that is based on recent U-Pb dating of zircons (Shi et al.[36]).

13. CG Spermatophytes | MRCA: *Ginkgo-Capsella*| 308.14 Ma – 365.629 Ma.
Fossil taxon and specimen. *Cordaites iowensis* [UIC 12,233: University of Illinois at Chicago; OUPH 9616- 9742: Ohio University Paleobotanical Herbarium, Department of Botany, Ohio University, Athens, Ohio] from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa[37].
Phylogenetic justification. Clarke *et al.*[1] identify cordaitean coniferophytes as the oldest records of the crown group of the spermatophyte clade. The oldest whole plant reconstruction is *Cordaites iowensis* from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa[37].
Minimum age. 308.14 Ma.
Soft maximum age. 365.629 Ma.
Age justification. Janousek and Pope[38] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus*[39]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone[40]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone[19,40]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al.[19], yielding a minimum constraint of 308.14 Ma.

The soft maximum constraint follows Clarke et al.[1] who based theirs on the first records of seeds in the form of preovules that satisfy the criteria of the seed habit, which occur in the Upper Fammenian (Late Devonian) VCo Spore Biozone[41], a well documented example of which being *Elkinsia polymorpha*[42]; *E. polymorpha* has been recovered from the Hampshire Formation, West Virginia, from which the palynomorphs *Grandispora cornuta, Retispora macroreticulata, Retusotriletes phillipsii* and *Rugospora radiata* have been reported[43], which substantiate assignment to the VCo Biozone[44]. The VCo biozone is not directly dated but its base falls within the *Palmatolepis trachytera* conodont biozone[45], the base of which is dated to 364.19 Ma ± 1.439 Myr[12], yielding a soft maximum constraint on the divergence of crown Spermatophyta at 365.629 Ma.

14. CG Acrogymnosperms | MRCA: *Ginkgo-Pinus* | 308.14 Ma – 365.629 Ma.
Fossil taxon and specimen. *Cordaites iowensis* [UM4616: University of Michigan and Illinois Geological Survey, Ann Arbor MI, USA] from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa, USA[37].
Phylogenetic justification. Clarke *et al.*[1] identify cordaitean coniferophytes as the oldest records of the *Ginkgo-Pinus* clade, the oldest whole plant reconstruction of which is *Cordaites iowensis* from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa[37].
Minimum age. 308.14 Ma.
Soft Maximum age: 365.629 Ma.
Age justification. Janousek and Pope[38] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola

Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus*[39]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone[40]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone[19,40]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al.[19], yielding a minimum age constraint of 308.14 Ma. A soft maximum is based upon the first appearance of seeds in the form of preovules which are attributable to the spermatophyte stem, the oldest interpretation of which is 365.629 Ma (see Spermatophyta).
Discussion. Zanne et al.[46] derive a minimum constraint from *Emporia lockardii* at 290.0 Ma which they recognize as a member of crown-Acrogymnospermae within a phylogenetic concept of the group in which, as here, cycads and *Ginkgo* comprise a clade.

15. MRCA: *Ginkgo-Cycas* | 264.7 Ma – 365.629 Ma.
Fossil taxon and specimen. *Crossozamia chinensis* [GP0027: Beijing Graduate School, China Institute of
Mining, Beijing, China], Lower Shihhotse Formation at Simugedong, Dongshan (East Hills), Taiyuan, north China[47].
Phylogenetic justification. Nagalingum et al.[48] identify *Crossozamia* as the oldest record of the *Cycas* lineage, based on megasporophylls that exhibit similarity to extant *Cycas*[49]. They argue against the interpretation of *Crossozamia* as the sister lineage of *Cycas* based on the presence of an estipulate leaf base and a terminal pinna found in the seedlings[49], instead favouring its assignment to the cycad stem. The arguments presented clearly raise doubts about the assignment of *Crossozamia* to crown-cycads, however, they do not provide definitive evidence of its exclusion from this clade and so *Crossozamia* may more appropriately be assigned to the cycad total group (i.e. we cannot discriminate between a stem or crown-cycad affinity based on the available evidence). In either instance, *Crossozamia* is the oldest record of the minimal clade comprised of *Gingko* and *Cycas*.
Minimum age. 264.7 Ma.
Soft Maximum age. 365.629 Ma.
Age justification. The Lower Shihhotse Formation at Simugedong, Dongshan (East Hills), Taiyuan, north China[47] has been established biostratigraphically as Roadian-Wordian (middle Permian)[50] and, thus a minimum age constraint can be established on the Wordian-Capitanian Boundary which has been dated to 265.1 Ma ± 0.4 Myr[51]. Thus, the minimum age constraint on the *Cycas-Ginkgo* clade is 264.7 Ma. A soft maximum is based upon the first appearance of seeds in the form of preovules which are attributable to the spermatophyte stem, the oldest interpretation of which is 365.629 Ma (see Spermatophyta).

16. CG Conifers | MRCA: Pinus-Cunninghamia | 147 Ma - 312.38 Ma.
Fossil taxon and specimen. *Araucaria mirabilis* [NHM V. 30953: Natural History Museum, London, UK], represented by cones, from Cerro Cuadrado petrified forest, La Matilde Formation, Patagonia, Argentina[52-55].
Phylogenetic justification. These fossils possess a 'vascular plexus' at the ovule base, ovuliferous scale vascularization, two vascular strands to the conescale complex and an embryo with two

cotyledons, all characters established to distinguish *Araucaria* section *Bunya* of the Araucariaceae[54,56], to which only extant *Araucaria bidwillii* belongs.

Minimum age. 147 Ma.

Soft Maximum age: 312.38 Ma.

Age justification. The age of La Matilde Formation is poorly constrained as the stratigraphy is complex, although the volcanic deposits do allow radiometric dating. La Matilde Formation is overlain by volcanics dated to 157 Ma ± 10 Myr[57], and thus the minimum constraint on the divergence of crown Cupressophyta, total group Cupressophyta and crown Coniferae is 147 Ma. A soft maximum constraint can be based on *Cordaites iowensis*, a cordaitean coniferophyte from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa[37], is the oldest whole plant reconstruction for Coniferae. Janousek and Pope[23] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus*[24]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone[40]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone[19,40]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al.[19], yielding a soft maximum of 312.38 Ma.

Discussion. This is the fundamental divergence of Coniferae into Cupressophyta, Gnetales and Pinaceae. The oldest secure records of the gnepine total group occur within the Yixian Formation of Liaoning, China, the minimum age of which is 121.8 Ma (see[1]). The oldest possible records of Cupressophyta total group include Triassic *Rissikia media* (Townrow, 1967) but it lacks the Podocarpaceae diagnostic feature of one ovule per cone scale, instead possessing two[1]. Other Triassic-Jurassic records are equally problematic[58-60].

17. CG Gnetales | MRCA: *Gnetum-Welwitschia* | 119.6 Ma – 312.38 Ma.

Fossil taxon and specimen. *Eoantha zherikhinii* [Repository of the Institute of Biology and Pedology, Vladivostok, Russia], from the Zaza Formation at the Baisa locality in the upper reaches of the Vitim River in Lake Baikal[61].

Phylogenetic justification.

Minimum age. 119.6 Ma.

Soft Maximum age. 312.38 Ma.

Age justification. The Zaza Formation can be correlated with the Turga Formation, also of Transbaikalia based principally on common elements of their floral assemblages, including *Asteropollis asteroids*, *Dicotylophyllum pusilum*, *Baisa hirsuita, Podozamites, Schizolepis, Pseudolarix, Phoenicopsis, Czekanowskia rigida* and *Sphenobaiera*[61-64]. The age of the Turga flora and Formation is based on the chronological distribution of *Asteropollis* type pollen, but correlation with the Yixian Formation of China is also supported strongly[62], allowing for refinement of the *Asteropollis*-derived ages. Correlation between Turga and Yixian is based on similarities in the floral assemblages of these two formations, with the shared presence of the species *Baisa hirsuita, Botrychites reheensis, Neozamites verchojanensis, Pityolepis pseudotsugaoides, Brachyphyllum longispicum, Scarbugia hilii, Ephedrites chenii, Carpolithus multiseminalis, Carpolithus pachythelis, Schizolepis, Baiera, Coniopteris, Ginkoites, Pityocladus* , *Pityospermum* and *Elatocladus*[61,62,65]. The shared presence of *Asteropollis asteroides* in Turga and Zaza

can be used to constrain their age. The last appearance of *Asteropollis* pollen is in Antarctica[66] and is dated to the end-Campanian, at the latest 72.1 Ma ± 0.2[67]. This minimum may be constrained further based on the correlation of the Zaza Formation through the Turga Formation to the Yixian Formation. The main fossil bearing beds in the Yixian Formation have been recently dated and may be as old as 129.2 Ma[68], however, in the absence of knowledge of the position of the fossils within the stratigraphy, relative to the sources of the absolute dates, a minimum age constraint can be derived from the Jiufontang Formation which overlies it. $^{40}Ar/^{39}Ar$ dating of a number of samples from the Jiufontang Formation has yielded an age of 120.3 ± 0.7 Ma for the volcanic tuffs[69], establishing a minimum constraint of 119.6 Ma for the age of the Yixian, Formation and, thus ultimately the Zaza Formation.

A soft maximum constraint can be based on *Cordaites iowensis*, a cordaitean coniferophyte from the Laddsdale Coals (Cherokee Group, Desmoinesian) near What Cheer, Iowa[37], is the oldest whole plant reconstruction for Coniferae. Janousek and Pope[38] argue that the Laddsdale Coal is equivalent to the Bluejacket Coal of Oklahoma, which occurs as part of the Bluejacket Sandstone Member, underlying the Inola Limestone, part of the Inola Cyclothem of the Krebs subgroup of the Cherokee Group, characterized by the occurrence of the conodonts *Idiognathodus amplificus, Idiognathodus podolskensis and Neognathodus asymmetricus*[39]. The Inola cyclothem falls fully within the *Idiognathodus amplificus/ Idiognathodus obliquus* biozone[40]. This is indicative of the *Neognathodus medexultimus-Streptognathodus concinnus* (Pc10) biozone, certainly older than the *Neognathodus roundyi – Streptognathodus cancellosus* (Pc11) biozone[19,40]. The base of Pc10 is bracketed by an older age constraint of 312.01 Ma ± 0.37 Myr and the base of Pc11 is bracketed by a younger age constraint of 308.5 Ma ± 0.36 Myr in the Composite Standard of Davydov et al.[19], yielding a soft maximum of 312.38 Ma.

18. CG Angiosperms | MRCA: *Amborella-Austrobuxus* | 125.9 Ma – 247.3 Ma.

Fossil taxon and specimen. Tricolpate pollen grain [BRN 126] from the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight[70].

Phylogenetic justification. Following Clarke at al.[1], our minimum age constraint is based on the earliest occurrences Fischer's rule tricolpate pollen, and knowledge of the distribution of tricoplate pollen across the phylogeny of angiosperms[71].

Minimum age. 125.9 Ma.

Soft maximum age. 247.3 Ma.

Age justification. Following Clarke at al.[1], the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight[70] occurs within the M1n polarity chron at the top of the Barremian, dated as 126.3 Ma ± 0.4 Myr[67]. The soft maximum age constraint is based on sediments devoid of angiosperm-like pollen below their first report in the Middle Triassic, thus, the base of the Anisian, dated to 247.1 Ma ± 0.2 Myr[35], thus, 247.3 Ma.

Discussion. The recently described *Euanthus panii*[72], *Juraherba bodae*[73] and *Yuhania dahugouensis*[74] from the Jiulongshan Formation were considered but not assigned. At the current stage, the age of the formation appears to be still not fully settled despite most experts agree on a middle Jurassic age (see[73,74]), whereas the assignment to extant lineages also required further investigation using phylogenetic approaches to confirm the proposed relationships of *Juraherba* to Hydatellaceae - which are the sister to the remaining Nymphaeales lineage and *Yuhania* to monocots.

19. SG Nympheales | MRCA: *Nymphaea-Kadsura* | 110.87 Ma.

Fossil taxon and specimen. *Pluricarpellatia peltata* [MB.Pb. 2000/80: Museum of Natural History, Berlin, Germany], from the Crato Formation of Brazil[75]
Phylogenetic justification. *Pluricarpellatia peltata* has been considered phylogenetically and resolved as members of the lineage leading to *Cabomba* after it diverged from *Nymphaea*[76].
Minimum age. 110.87 Ma.
Age justification. Clarke et al.[1] argued that the age of the Crato Formation could not be constrained to being definitively older than Albian based on pollen[77], ostraco[78] and dinoflagellate[79] biostratigraphy and, in the absence of further evidence, established a minimum constraint on the Albian-Cenomanian boundary. Massoni et al.[80] argued for an Aptian age for the Crato Formation based on evidence from Heimhofer and Hochuli[79] but, unfortunately, these authors do not present evidence that can discriminate against a possible early Albian age for the Crato Formation, as acknowledged by Mohr et al.[81]. While the evidence suggests, at worst, an early Albian age for the Crato Formation, it is possible to derive a minimum age interpretation for the Formation based on the Early-Middle Albian Boundary, which coincides approximately with the base of the *Douvilleiceras mammillatum* ammonite biozone, dated to 110.87 Ma[67].
Discussion. Magallon et al.[23] derive a minimum constraint from *Monetianthus mirus* which they recognize as a representative of the Nymphaeaceae stem lineage and, thus, use it as the basis of a minimum constraint on the age of total-group Nymphaceaceae at 125 Ma. However, Clarke et al.[1] demonstrated that the minimum age of the host deposit, Vale de Água, Portugal[82,83] is 93.9 Ma[67]. However, there are other, potentially older records of Nymphaeaceae and, more specifically, the crown clade circumscribed by *Nymphaea-Cabomba*. Clarke et al.[1] identified much older, but more equivocal records, as well as the oldest unequivocal records, viz. *Pluricarpellatia peltata* from the Crato Formation of Brazil[75] and *Scutifolium jordanicum* from the Jarash Formation (Kurnub Group) of Jordan[76], both of which have been considered phylogenetically and resolved as members of the lineage leading to *Cabomba* after it diverged from *Nymphaea*[76]. *Scutifolium jordanicum* was used to establish a minimum age for crown-Nymphaeales at 105 Ma by Smith et al.[84], and for total-group Cabombaceae at 105 Ma by Zanne et al.[46]. The Jarash Formation can be dated minimally to 95 Ma (96.1 Ma ± 1.1 Myr in[85], but the Crato Formation is older .

20. SG Austrobaileyales | MRCA: *Kadsura* - *Capsella*| 107.59 Ma.
Fossil taxon and specimen. *Anacostia virginiensis* [PP44151] from the Puddledock locality, Tarmac Lone Star Industries sand and gravel pit, Virginia USA[86].
Phylogenetic justification. Originally ascribed as an early magnoliid or monocot[86], Doyle et al.[87] resolved through phylogenetic analysis that *Anacostia* belongs within the Austrobaileyales based on the presence of several synapomorphies including a sclerotic mesotesta, palisade exotesta and basal ovule position.
Minimum age. 107.59 Ma.
Age justification. Massoni et al.[80] reason that the sediments in the Puddledock Locality are definitively early Albian based on the presence of reticulate tricolpate pollen and *Clavatipollenites rotundus* (aff. *Retimonocolpites dividuus*[88]) but not striate tricolpates, which occur later in the early Albian. Therefore, they constrain minimally the age of the *A. virginiensis* by the Middle-late Albian boundary, which coincides with the base of the *Diploceras cristatum* biozone which has been dated to 107.59 Ma[67].
Discussion. *Anacostia*, reportedly from the early and middle Albian of Buarcos, Famalicão, and Vale de Agua (Portugal), Puddledock (Virginia, USA), and Kenilworth (Maryland, USA) was recognized as

the oldest fossil record of the Austrobaileyales[89,90]. Doyle and Endress[90] *identified Anacostia portugallica* and *A. teixeirae* as early Albian and, therefore the oldest species belonging to this lineage. However, the minimum age interpretation of these localities the Figueira da Foz Formation cannot be constrained minimally to more than 92.8 Ma (see above). However, the minimum age constraint on *A. virginiensis* from the Puddledock Locality is older.

21. CG Mesangiosperms | MRCA: *Sarcandra* – *Capsella* | 125.9 Ma – 247.3 Ma.
Fossil taxon and specimen. Tricolpate pollen grain [BRN 126] from the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight.[70].
Phylogenetic justification. Following Clarke at al.[1], our minimum age constraint is based on the earliest occurrences Fischer's rule tricolpate pollen, and knowledge of the distribution of tricoplate pollen across the phylogeny of angiosperms[71].
Minimum age. 125.9 Ma.
Soft maximum age. 247.3 Ma.
Age justification. Following Clarke at al.[1], the Cowleaze Chine Member of the Vectis Formation of the Isle of Wight[70] occurs within the M1n polarity chron at the top of the Barremian, dated as 126.3 Ma ± 0.4 Myr[67]. The soft maximum age constraint is based on sediments devoid of angiosperm-like pollen below their first report in the Middle Triassic, thus, the base of the Anisian, dated to 247.1 Ma ± 0.2 Myr[9], thus, 247.3 Ma.

22. CG Magnoliales | MRCA: *Liriodendron - Persea* | 110.87 Ma.
Fossil taxon and specimen. *Endressinia brasiliana* [MB. PB. 2001/1455: Museum of Natural History, Institute of Paleontology, Berlin, Germany], from the Crato Formation of Brazil[90].
Phylogenetic justification. Masson et al. identify both *Schenkeriphyllum glanduliferum* and *Endressinia brasiliana*, both from the Crato Formation of Brazil[90,91], as the oldest records of crown Magnoliineae, the sister clade of Myristicaceae [92], based on the phylogenetic analyses[89,90,91].
Minimum age. 110.87 Ma.
Age justification. Clarke et al.[1] argued that the age of the Crato Formation could not be constrained to being definitively older than Albian based on pollen[77], ostracod[78], and dinoflagellate[79] biostratigraphy and, in the absence of further evidence, established a minimum constraint on the Albian-Cenomanian boundary. Massoni et al.[80] argued for an Aptian age for the Crato Formation based on evidence from Heimhofer and Hochuli[79] but, unfortunately, these authors do not present evidence that can discriminate against a possible early Albian age for the Crato Formation, as acknowledged by Mohr et al.[92]. While the evidence suggests at worst, an early Albian age for the Crato Formation, and so it is possible to derive a minimum age interpretation for the Formation based on the Early-Middle Albian Boundary, which coincides approximately with the base of the *Douvilleiceras mammillatum* ammonite biozone, dated to 110.87 Ma[67].
Discussion. *Archaeanthus linnenbergii* was recognized as a further putative stem group Magnoliaceae but it is younger than *Endressinia*[80,89].

23. SG Saururaceae| MRCA: *Saruma-Houttuynia* | 44.3 Ma.
Fossil taxon and specimen. *Saururus tuckerae* [UAPC P1631 Bbot a: University of Alberta (Edmonton) Paleobotanical Collections] from the Middle Eocene Princeton Chert, British Columbia, Canada.
Phylogenetic justification. Massoni et al.[80] follow Smith and Stockey[94] in identifying *Saururus tuckerae* as the oldest record of total group *Saururus*. Based on tens of flowers and a partial inflorescence, the

flower structure and pollen are characteristic of Saururaceae (Piperales), and phylogenetic analyses resolved *S. tuckerae* as the sister clade to extant *Saururus*[94].

Minimum age. 44.3 Ma.

Age justification. The Princeton Chert is part of the Allenby Formation which has been the subject of a number of absolute dating studies yielding age estimates of 48 Ma ± 2 Myr[95,96], between 47 Ma ± 2 Myr and 50 Ma ± 2 Myr[97], between 46.2 Ma ± 1.9 Myr and 49.4 Ma ± 2 Myr[98], and 52.08 Ma ± 0.12 Myr[99] for the Allenby Formation. We follow Massoni et al.[80] in basing our minimum constraint based on the youngest age Interpretation of the youngest radiometric age estimate, viz. 44.3 Ma

24. CG Monocots | MRCA: *Acorus-Oryza* | 112.6 Ma.

Fossil taxon. The earliest records of *Liliacidites* occur at the Trent's Reach Locality of the Potomac Group, attributable to the Albian Zone I[100].

Phylogenetic justification. Doyle et al.[87] identified pollen referred to the genus *Liliacidites* (but not Similipollis) as represesentative of the monocot stem, making it the oldest secure record of the monocot total group (see[89]).

Minimum age. 112.6 Ma.

Age justification. In the absence of further stratigraphic constraint, these earliest records of *Liliacidites* can be constrained in age by the Aptian-Albian Boundary, dated to 113.0 ma ± 0.4 Myr, thus, 112.6 Ma.

Discussion. Doyle et al.[87] highlight that, despite decades of sampling of the Hauterivian and Barremian of England, no clear representatives of *Liliacidites* pollen have been recovered[101], perhaps implying that the earliest records from the Albian are a close approximation of their antiquity. Because of the position of monocots in our molecular tree we consider *Liliacidites* to be nested within monocots, and use it to calibrate the monocot crown node.

25. CG Coryphoideae | MRCA: *Sabal-Oryza* | 83.41 Ma.

Fossil taxon and specimen. *Sabalites carolinensis* [PAL 175717/P 38208: Smithsonian Museum of Natural History; Washington DC, USA] described from the Middendorf Arkose Member of Black Creek Formation near Langley, Aiken County, South Carolina[102].

Phylogenetic justification. The phylogenetic relationships of this fossil have been discussed in Hertweck et al.[103] and Iles et al.[104].

Minimum age. 83.41 Ma

Age justification. Berry's view that the Middendorf was merely a distinct facies within the Black Creek Formation, rather than a stratigraphically distinct unit, has been rejected. Sohl and Owens[112] subdivided the Upper Cretaceous of Carolina coastal plain into three lithostratigraphic units, the Middendorf, Black Creek and Peedee Formations, raised the Black Creek to group status and subdivided this into three unconformity-bound formations, viz. in stratigraphic sequence, the Tar Heel, Bladen and Donoho Creek formations. Evidently, *Sabalites carolinensis* was recovered from what is now recognized as the Middendorf Formation, and a minimum age constraint can be established on the boundary between the Middendorf and Tar Heel Formations. The Middendorf is commonly considered Santonian in age, however, little material evidence has been presented in support of this, in part a consequent of the complex history of stratigraphic divisions at outcrop, in subsurface and offshore[105]. Habib and Miller[106] established an age 'not younger than Campanian' on the basis of dinoflagellate biostratigraphy, but following the stratigraphic scheme outlined Campbell and Grohn[105], the Middendorf is older that the Shepherd Grove Formation and, therefore, following the

stratigraphy of Christopher and Prowell[107], must be no younger than Santonian. Thus, we may established a minimum age constraint on the *Sabalites carolinesis* based on the Santonian-Campanian Boundary, coincident with the base of the *Scaphites leei III* Zone, dated to 83.64 Ma ± 0.23 Myr[67], thus, 83.41 Ma.

26. SG Musaceae | MRCA: *Musa-Oryza* | 74.6 Ma.
Fossil taxon and specimen. *Spirematospermum chandlerae* has been described from isolated seeds and groups of seeds from the Neuse River locality, Black Creek Formation, southwest of Goldsboro, Wayne County, North Carolina, USA.
Phylogenetic justification. The phylogenetic relationships of this fossil have discussed in previous studies[103].
Minimum age. 74.6 Ma.
Age justification. Reputedly Late Cretaceous (Early Campanian) in age[108], the Black Creek Formation has been assigned to the *Exogyra ponderosa* Biozone which occurs beneath the *Didymoceras cheyennense Tethyan ammonoid biozone*[107], the base of which is dated to 74.6 Ma[67].

27. SG Dioscoreales | MRCA: *Dioscorea-Colchicum* | 85.8 Ma.
Fossil taxon and specimen. *Mabelia connatifila* [CUPC 1255: L. H. Bailey Hortorium Paleobotanical Collection,
Cornell University, Ithaca NY, USA] from the South Amboy Fire Clay Member of the Raritan Formation at the Old Crossman clay pit in Sayreville, New Jersey, USA[109].
Phylogenetic justification. The phylogenetic assignment is based on the phylogenetic hypothesis reconstructed by Gandolfo et al.[109].
Minimum age. 85.8 Ma.
Age justification. Clarke et al.[1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however, Massoni et al.[80] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

28. SG Oryzeae | MRCA *Oryza* – *Brachypodium* | 65.98 Ma.
Fossil taxon and specimen. *Changii indicum* [Slide 13160, Q-14-3, Birbal Sahni I. Palaeobotany, Lucknow, India] from the Maastrichtian-Danian Deccan beds of India
Phylogenetic justification. The phylogenetic relationships of this fossil have discussed in previous studies[103].
Minimum age. 65.98 Ma.
Age justification. We follow Iles *et al.*[103] and their recommendation of the radiometric and magnetostrategraphic dating of the Deccan beds of India by Courtillot and Ren[110] and the presence of dinosaur coprolites to be latest Maastrichtian, updated following Ogg & Hinnov[67]

29. CG Eudicots | MRCA: *Escholzia-Capsella*| 119.6 Ma.
Fossil taxon and specimen. *Hyrcantha decussata* [NJU-DES02001: Geological Institute, Chinese Academy of Sciences, Beijing], from the lower part of the Yixian Formation, Jehol Group, Liaoning Province, China[111].
Phylogenetic justification. Similar to *Leefrutcus* from the Yixian formation of the Lower Cretaceous of China, *Hycantha* is considered to be a stem group representative of the Ranunculales[112].

Minimum age. 119.6 Ma.
Age justification. The main fossil bearing beds have been dated and may be as old as 129.2 Ma[68], however, in the absence of knowledge of the position of the fossils within the stratigraphy, relative to the sources of the absolute dates, a minimum age constraint can be derived from the Jiufontang Formation which overlies it. $^{40}Ar/^{39}Ar$ dating of a number of samples from the Jiufontang Formation has yielded an age of 120.3 ± 0.7 Ma for the volcanic tuffs[69], establishing a minimum constraint of 119.6 Ma.

30. CG Ericales core | MRCA: *Diospyros-Inula* | 85.8 Ma.
Fossil taxon and specimen. *Paleoenkianthus sayrevillensis* [CUPC 1100: L. H. Bailey Hortorium, Cornell University, Ithaca NY, USA] from the South Amboy Fire Clay of the Raritan Formation, of which outcrops are exposed in the Old Crossman Clay Pit in Sayreville, New Jersey.
Phylogenetic justification. The phylogenetic relationships of this fossil has been tested based on morphological evidence[113].
Minimum age. 85.8 Ma
Age justification. Clarke et al.[1] argued that a minimum constraint on the age of this deposit could be established from Santonian-Campanian Boundary, however, Massoni et al.[80] argue that a tighter correlation can be established with better rocks attributable to the CC13-14 Nannofossil zones in South Carolina, indicating a minimum age of 86.3 Ma ± 0.5 Myr, thus, 85.8 Ma.

31. SG Asteraceae minus *Bernadesia* | MRCA: *Tanacetum - Inula* | 41.5 Ma.
Fossil taxon and specimen. *Tubulifloridites antipodica* from onshore deposits taken from a paleochannel at Koingnaas, on the west coast of South Africa.
Phylogenetic justification. The newly described *Tubulifloridites lilliei* type A predates this estimates with an age of 76 – 66 Ma[114] , however the assignment of this fossil and its affinity with Asteraceae remains controversial[115]. The placement of the pollen fossils of *T. antipodica* within Asteraceae minus *Bernadesia* is deemed reliable[116].
Minimum age. 41.5 Ma.
Age justification. These occurrences are, described to occur alongside the planktic forams *Globigerinatheka index* and *Turborotalia centralis*[117]. *Globigerinatheka index* is known to range from 42.9 - 34.3 Ma[118], but *Turborotalia centralis* is a junior synonym of *Turborotalia pomeroli*, which is known to range from 42.4-41.5 Ma[118]. Thus, the minimum age constraint on *Tubulifloridites antipodica* is 41.5 Ma.

32. SG Myrtales | MRCA: *Eucalyptus-Capsella* | 83.3 Ma.
Fossil taxon and specimen. *Esqueiria futabensis* [PP45419: Field Museum, Chicago IL, USA] from two levels in the Futaba Group exposed in Fukushima Prefecture. northeastern Honshu, Japan[119].
Phylogenetic justification. The phylogenetic relationships have been established by several authors[120].
Minimum age. 83.3 Ma.
Age justification. One locality, considered Coniacian, occurs in the Asamigawa Member of the Ashizawa Formation, on a tributary of the Kitaba River in Kamikitaba, Hirono-machi. Unfortunately, no material evidence has been presented to substantiate this age assignment (Takahashi et al.[119], among others, merely cite the presence of unspecified Coniacian ammonites). The second locality is in the middle part of the Tamayama Formation, on the Kohisa River, Kohisa, Ouhisa machi, northeast of lwaki City. The Asamigawa Formation is the lowermost formation in the Futaba Group, and is

overlain by the Kasamatsu Formation, in turn overlain by the Tamayama Formation. The age of the Tamayama Formation is substantiated on the presence of *Inoceramus amakusensis*[119], which is restricted to the Santonian[121]. Thus, a minimum age constraint may be established on the Santonian-Campanian Boundary, dated as 83.6 Ma ± 0.3 Myr[67], thus, 83.3 Ma.

33. SG Sapindales | MRCA: *Citrus-Capsella* | 59.24 Ma.
Fossil taxon and specimen. *Dipteronia brownii* [UF 15740E-23086: Florida Museum of Natural History, Gainesville FL, USA] from the Paleocene Fort Union Formation at Hell's Half Acre, Wyoming[122].
Phylogenetic justification. This fossil is assigned to the extant genus *Dipteronia* which belongs to the subfamily Hippocantanoides of the family Sapindaceae. The extant genus is considered a Tertiary relict having two extant species endemic to China[123,124]. Being a possible stem group representative of the extant genus nested in the Sapindales provided the framework for this assignment.
Minimum age. 59.24 Ma.
Age justification. *Dipteronia brownii* occurs within the P4 Pollen Zone in the type section of Nichols and Ott[125], which falls fully within Magnetic Anomaly Zone C26r[126], the end of which is dated to 59.24 Ma in the combined age model of Vandenberghe et al.[127].

34. SG *Salicaceae* | MRCA: *Linum-Populus* | 48.57 Ma.
Fossil taxon and specimen. *Pseudosalix handleyi* [UMNH PB-1: Utah Museum of Natural History, Salt Lake City, USA] from lacustrine shales of the Parachute Creek Member of the Green River Formation in the vicinity of Bonanza, Utah, USA[128].
Phylogenetic justification. Our node assignment follows the currently accepted interpretation of the fossil record of Salicaceae[129].
Minimum age. 48.57 Ma.
Age justification. The Parachute Creek Member reaches into C22n magnetozone[130], the minimum age of which can be established from the base of the succeeding C21r, dated to 48.57 Ma in the combined age model of Vandenberghe et al.[127]

1. Clarke, J., Donoghue, P. C. J. & Warnock, R. C. M. Establishing a timescale for plant evolution. *New Phytologist* **192**, 266-301 (2011).
2. Steemans, P. *et al.* Origin and Radiation of the Earliest Vascular Land Plants. *Science* **324**, 353-353 (2009).
3. Strother, P. K., Wood, G. D., Taylor, W. A. & Beck, J. H. Middle Cambriancryptospores and the origin of land plants. *Memoirs of the Association of Australasian Palaeontologists* **24**, 99–113 (2004).
4. Cooper, R. A. & Sadler, P. M. in *Geologic timescale 2012* (eds Felix M. Gradstein, James G. Ogg, Mark Schmitz, & Gabbi Ogg) 489-523 (Elsevier, 2012).
5. Peng, S., Babcock, L. E. & Cooper, R. A. in *Geologic timescale 2012* (eds Felix M. Gradstein, James G. Ogg, Mark Schmitz, & Gabbi Ogg) 437-488 (Elsevier, 2012).
6. Townrow, J.A. Two Triassic Bryophytes from South Africa. *South African Journal of Botany* **25**, 1-22 (1959).
7. Anderson, H.A. A review of the Bryophyta from the Upper Triassic Molteno Formation, Karoo Basin, South Africa. *Palaeontologica Africana* **19,** 21-30 (1976)

8. Anderson, J.M. & Anderson, H.M. The fossil content of the Upper Triassic Molteno Formation, South Africa. *Palaeontologica Africana* **25**, 39–59 (1984)
9. Ogg, J. G. in *The geologic time scale 2012* (eds F. M. Gradstein, J. G. Ogg, M. Schmitz, & G. Ogg) 681-730 (Elsevier, 2012).
10. Guo, C. Q., Edwards, D., Wu, P. C., Duckett, J. G., Hueber, F. M. & Li, C. S. *Riccardiothallus devonicus* gen. et sp. nov., the earliest simple thalloid liverwort from the Lower Devonian of Yunnan, China. *Review of Palaeobotany and Palynology* **176,** 35-40 (2012)
11. Hao, S. G., Gensel, P. G. & Wang, D. M. *Polythecophyton demissum*, gen.et sp. nov., a new plant from the Lower Devonian (Pragian) of Yunnan, China and its phytogeographic significance. *Review of Palaeobotany and Palynology* **116**, 55–71 (2001).
12. Becker, R. T., Gradstein, F. M. & Hammer, O. in *The geological timescale 2012* (eds F. M. Gradstein, J. G. Ogg, M. Schmitz, & G. Ogg) 559-601 (Elsevier, 2012).
13. Kenrick, P. & Crane, P. R. *The origin and early diversification of land plants: a cladistic study*. (Smithsonian Institution Press, 1997)
14. Zalasiewicz, J. A., Taylor, L., Rushton, A. W. A., Loydell, D. K., Rickards, R. B. & Williams, M. Graptolites in British stratigraphy. *Geological Magazine* **146**, 785-850 (2009).
15. Melchin, M. J., Sadler, P. M. & Cramer, B. D. in *Geologic timescale 2012* (eds Felix M. Gradstein, James G. Ogg, Mark Schmitz, & Gabbi Ogg) 525-558 (Elsevier, 2012).
16. De Souza, I. C. C., Branco, F. S. R. & Vargas, Y. L. Permian bryophytes of Western Gondwanaland from the Parana Basin in Brazil. *Palaeontology* **55,** 229-241 (2012).
17. Laenen, B. et al. (2014) Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nature Communications* **5,** 6134 (2014).
18. Santos, R. V., Souza, P. A., Alvarenga, C. J. S., Dantes, E. L., Pimentel, M. M., Oliveira, C. G. & Araújo, L. M. Shrimp U-Pb Zircon dating and Palinology of Bentonitic layers from Permian Irati Formation, Paraná Basin, Brazil. *Gondwana Research* **9**, 41– 65 (2006).
19. Davydov, V. I., Korn, D. & Schmitz, M. D. in *The geologic time scale 2012* (eds F. M. Gradstein, J. G. Ogg, M. Schmitz, & G. Ogg) 603-651 (Elsevier, 2012)
20. Kotyk, M. E., Basinger, J. F., Gensel, P. G. & de Freitas, T. A. Morphologically complex plant macrofossils from the Late Silurian of Arctic Canada. *American Journal of Botany* **89**, 1004-1013 (2002).
21. Wellman, C. H., Gensel, P. G. & Taylor, W. A. Spore wall ultrastructure in the early lycopsid *Leclercqia* (Protolepidodenrales) from the Lower Devonian of North America: evidence for a fundamental division in the lycopsids. *American Journal of Botany* **96**, 1849-1860 (2009).
22. Meyer-Berthaud, B., Fairon-Demaret, M., Steemans, P., Talent, J. & Gerrienne, P. (2003) The plant *Leclercqia* (Lycopsida) in Gondwana: implications for reconstructing Middle Devonian palaeogeography. *Geological Magazine* **140**, 119-130.
23. Magallon, S., Hilu, K. W. & Quandt, D. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* **100**, 556-573 (2013).
24. Bonamo, P. M. (1977) *Rellimia Thomsonii* (Progymnospermopsida) from Middle Devonian of New York State. *American Journal of Botany* **64**, 1272-1285.
25. Skog, J. E. & Banks, H. P. (1973) *Ibyka amphikoma*, gen et sp-n - new protoarticulate precursor from late Middle Devonian of New York State. *American Journal of Botany* **60**, 366-380.

26. Bartholomew, A. J. & Brett, C. E. (2007) Correlation of Middle Devonian Hamilton Group-equivalent strata in east-central North America: implications for eustasy, tectonics and faunal provinciality. *Geological Society, London, Special Publications* **278**, 105-131.
27. Johnson, J. G., Klapper, G. & Sandberg, C. A. (1985) Devonian eustatic fluctuations in Euramerica. *Geological Society of America Bulletin* **96**, 567.
28. Fisher, D. W., Isachsen, Y. W., Rickard, L. V., Broughton, J. G. & Offield, T. W. *Geologic map of New York*. (New York State Museum Sciece Service, Geological Survey, 1962).
29. Rickard, L. V. *Correlation of the Devonian rocks in New York State. Map and Chart Series 4*. (New York State Musuem Science Service Geological Survey, 1964).
30. Klapper, G. in *Devonian biostratigraphy of New York, Part I.* (eds W. A. Oliver, Jr. & G. Klapper) 57-68 (IUGS SDS, 1981).
31. Kirchgasser, W. T. (2000) Correlation of stage boundaries in the Appalachian Devonian, Eastern United States. *Courier Forschungsinstitut Senckenberg* **225**, 271-284.
32. Pšenička, J. & Bek, J. Cuticles and spores of *Senftenbergia plumosa* (Artis) Bek and Pšenička from the Carboniferous of Pilsen Basin, Bohemian Massif. *Review of Palaeobotany and Palynology* **125**, 299-312 (2003)
33. Bek, J. & Pšenička, J. *Senftenbergia plumosa* (Artis) emend and their microspores from the Carboniferous of the Kladno and Pilsen. *Review of Palaeobotany and Palynology*. **116,** 213-232 (2001).
34. Schmidt, A. R., Heinrichs, J. & Schneider, H. Burmese amber fossils bridge the gap in the Cretaceous record of polypod ferns. *Perspectives in Plant Ecology, Evolution and Systematics* **18**, 70-78 (2016)
35. Cruickshank, R. D. & Ko, K. Geology of an amber locality in the Hukawng valley, Northern Myanmar. *Journal of Asian Earth Sciences* **21**, 441–445 (2003).
36. Shi, G. H., Grimaldi, D. A., Harlow, G. E., Wang, J., Yang, M. C., Lei, W. Y., Li, Q. L. & Li, X. H. Age constraint on Burmese amber based on U-Pb dating of zircons. *Cretaceous Researc*h **37,** 155–163 (2012).
37. Trivett, M. L. Growth architecture, structure, and relationships of *Cordaixylon iowensis* nov comb (Cordaitales). *International Journal of Plant Sciences* **153**, 273-287 (2002)
38. Janousek, T. J. & Pope, J. P. Petrology, petrography and conodont biostratigraphy of the Laddsdale Coal interval, along Whitebreast Creek, Bauer, Iowa. *GSA North-Central Section, 48th Annual Meeting, Abstracts* **16-5** (2014)
39. Heckel, P. H. Pennsylvanian stratigraphy of Northern Midcontinent Shelf and biostratigraphic correlation of cyclothems. *Stratigraphy* **10**, 3-39 (2013)
40. Barrick, J. E., Lambert, L. L., Heckel, P. H., Rosscoe, S. J. & Boardman, D. R. Midcontinent Pennsylvanian conodont zonation. *Stratigraphy* **10**: 55-72 (2013).
41. Prestianni, C. Early diversification of seeds and seed-like structures. *Carnets De Geologie*, 33-38 (2003)
42. Rothwell, G. W., Scheckler, S. E. & Gillespie, W. H. *Elkinsia* gen nov, a late Devonian gymnospermn with cupulate ovules. *Botanical Gazette* **150**: 170-189 (1989).
43. Streel, M. & Scheckler, S. E. Miospore lateral distribution in upper Fammenian alluvial lagoonal to tidal facies from eastern United States and Belgium. *Review of Palaeobotany and Palynology* **64**, 315-324 (1990).

44. Streel, M., Higgs, K., Loboziak, S., Riegel, W. & Steemans, P. Spore stratigraphy and correlation with faunas and floras in the type marine Devonian of the Ardenne-Rhenish regions. *Review of Palaeobotany and Palynology* **50**, 211-229 (1987).
45. House, M. R. & Gradstein, F. M. in *A geologic timescale 2004* (eds F. M. Gradstein, J. G. Ogg, & A. G. Smith) 202-221 (Cambridge University Press, 2004).
46. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89-92, (2014).
47. Gao, Z. & Thomas, B. A. A review of fossil cycad megasporophylls, with new evidence of Crossozamia pomel and its associated leaves from the lower Permian of Taiyuan, China. *Review of Palaeobotany and Palynology* **60**, 205-223 (1989).
48. Nagalingum, N. S. *et al.* Recent synchronous radiation of a living fossil. *Science* **334**, 796-799 (2011).
49. Hermsen, E. J., Taylor, T. N., Taylor, E. L. & Stevenson, D. W. Cataphylls of the Middle Triassic cycad Antarcticycas schopfii and new insights into cycad evolution. *American Journal of Botany* **93**, 724-738 (2006).
50. Wang, J. Late Paleozoic macrofloral assemblages from Weibei Coalfield, with reference to vegetational change through the Late Paleozoic Ice-age in the North China Block. *International Journal of Coal Geology* **83**, 292-317 (2010).
51. Henderson, C. M., Gradstein, F. M. & Hammer, O. in *The geologic timescale 2012* (eds F. M. Gradstein, J. G. Ogg, M. Schmitz, & G. Ogg) 653-679 (Elsevier, 2012).
52. Wieland, G. W. The Cerro Cuadrado petrified forest. *Carnegie Institution of Washington Publication* **449**, 1-183 (1935).
53. Calder, M. G. A coniferous petrified forest in Patagonia. *Bulletin of the British Museum (Natural History): Geology* **2**, 99-138 (1953).
54. Stockey, R. A. Seeds and embryos of *Araucaria mirabilis*. *American Journal of Botany* **62**, 856-868 (1975).
55. Stockey, R. A. Reproductive biology of Cerro Cuadrado fossil conifers: Ontogeny and reproductive strategies in *Araucaria mirabilis* (Spegazzini) Windhausen. *Palaeontographica Abteilung B* **166**, 1-15 (1978).
56. Wilde, M. H. & Eames, A. J. The ovule and seed of *Araucaria badwillii* with discussion of the taxonomy of the genus. 1. Morphology. *Annals of Botany* **12**, 311-& (1948).
57. Spalleti, L., Iñiguez Rodríguez, A. M. & Masón, M. Edades radimétricas de piroclastitas y volvanitas del Grupo Bahía Laura, Gran Bajo de San Julián, Santa Cruz. *Revista de la Asociación Geológica Argentina* **37**, 483-485 (1982).
58. Florin, R. Evolution in cordaites and conifers. *Acta Horti Bergiani* **15**, 285-388 (1951).
59. Yao, X. L., Taylor, T. N. & Taylor, E. L. A taxodiaceous seed cone from the Triassic of Antarctica. *American Journal of Botany* **84**, 343-354 (1997).
60. Axsmith, B. J., Taylor, T. N. & Taylor, E. L. Anatomically preserved leaves of the conifer Notophytum krauselii (Podocarpaceae) from the Triassic of Antarctica. *American Journal of Botany* **85**, 704-713 (1998).
61. Krassilov, V. A. New floral structure from the Lower Cretaceous of Lake Baikal Area. *Review of Palaeobotany and Palynology* **47**, 9-16 (1986)
62. Godefroit, P. *Bernissart dinosaurs and Early Cretaceous terrestrial ecosystems*. (Indiana University Press, 2012).

63. Vakhrameev, V. & Kotova, I. Ancient angiosperms and accompanying plants from the Lower Cretaceous of Transbaikalia. *Paleontological Journal* **4**, 487-495 (1977).
64. Vakhrameev, V. *Jurassic and Cretaceous floras and climates of the Earth*. (Cambridge University Press, 1991).
65. Chen, P. *et al.* Jianshangou Bed of the Yixian Formation in West Liaoning, China. *Science in China Series D: Earth Sciences* **48**, 298-312 (2005).
66. Dettmann, M. E. & Thomson, M. R. A. Cretaceous palynomorphs from the James-Ross Island area, Antarctica - a pilot-study. *British Antarctic Survey Bulletin* **77**, 13-59 (1987).
67. Ogg, J. G. & Hinnov, L. A. in *The geologic time scale 2012* Vol. 2 (eds F. M. Gradstein, J. G. Ogg, M. Schmitz, & G. Ogg) 793-853 (Elsevier, 2012).
68. Chang, S.-C., Zhang, H., Hemming, S. R., Mesko, G. T. & Fang, Y. Chronological evidence for extension of the Jehol Biota into Southern China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **344**–**345**, 1-5 (2012).
69. He, H. Y. *et al.* Timing of the Jiufotang Formation (Jehol Group) in Liaoning, northeastern China, and its implications. *Geophysical Research Letters* **31**, 1-4 (2004).
70. Hughes, N. F. & McDougall, A. B. Barremian-Aptian angiospermid pollen records from southern England. *Review of Palaeobotany and Palynology* **65**, 145-151 (1990).
71. Judd, W. S. & Olmstead, R. G. A survey of tricolpate (eudicot) phylogenetic relationships. *American Journal of Botany* **91**, 1627-1644 (2004).
72. Liu, Z. J. & Wang, X. A perfect flower from the Jurassic of China. *Hist Biol* **28**, 707-719 (2016).
73. Han, G. *et al.* A whole plant herbaceous angiosperm from the Middle Jurassic of China. *Acta Geologica Sinica* **90**, 19-29 (2016).
74. Liu, Z.-J. & Wang, X. Yuhania: a unique angiosperm from the Middle Jurassic of Inner Mongolia, China. *Historical Biology*, 1-11 (2016).
75. Mohr, B. A. R., Bernardes-De-Oliveira, M. E. C. & Taylor, D. W. Pluricarpellatia, a nymphaealean angiosperm from the Lower Cretaceous of northern Gondwana (Crato Formation, Brazil). *Taxon* **57**, 1147-1158 (2008).
76. Taylor, D. W., Brenner, G. J. & Basha, S. H. Scutifolium jordanicum gen. et sp nov (Cabombaceae), an aquatic fossil plant from the Lower Cretaceous of Jordan, and the relationships of related leaf fossils to living genera. *American Journal of Botany* **95**, 340-352 (2008).
77. Batten, D. J. in *The Crato fossil beds of Brazil - window into an ancient world* (eds D. M. Martill, G. Bechly, & R. F. Loveridge) 566-573 (Cambridge University Press, 2007).
78. Martill, D. M. The age of the Cretaceous Santana Formation fossil Konservat Lagerstatten of north-east Brazil: a historical review and an appraisal of the biochronostratigraphic utility of its palaeobiota. *Cretaceous Research* **28**, 895-920 (2007).
79. Heimhofer, U. & Hochuli, P.-A. Early Cretaceous angiosperm pollen from a low-latitude succession (Araripe Basin, NE Brazil). *Review of Palaeobotany & Palynology* **161**, 105-126 (2010).
80. Massoni, J., Doyle, J. A. & Sauquet, H. Fossil calibration of Magnoliidae, an ancient lineage of angiosperms. *Palaeontologica Electronica* (2014).
81. Mohr, B. A. R., Coiffard, C. & Bernardes-de-Oliveira, M. E. C. Schenkeriphyllum glanduliferum, a new magnolialean angiosperm from the Early Cretaceous of Northern Gondwana and its relationships to fossil and modern Magnoliales. *Review of Palaeobotany and Palynology* **189**, 57-72 (2013).

82. Friis, E. M., Pedersen, K. R. & Crane, P. R. Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* **410**, 357-360 (2001).
83. Friis, E. M., Pedersen, K. R., Von Balthazar, M., Grimm, G. W. & Crane, P. R. *Monetianthus mirus* gen. et sp. nov., a nymphaealean flowers from the Early Cretaceous of Portugal. *International Journal of Plant Science* **170**, 1086-1101 (2009).
84. Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences* **107**, 5897-5902 (2010).
85. Amireh, B. S., Jarrar, G., Henjes-Kunst, F. & Schneider, W. K-Ar dating, X-ray diffractometry, optical and scanning electron microscopy of glauconites from the Early Cretaceous Kurnub. *Geological Journal* **33**, 49-65 (1998).
86. Friis, E.M., Crane, P.R. & Pedersen, K.R. *Anacostia*, a new basal angiosperm from the Early Cretaceous of North America and Portugal with trichotomocolpate/monocolpate pollen. *Grana* **36**, 225-244 (1997).
87. Doyle, J. A., Endress, P. K. & Upchurch, G. R., Jr. Early Cretaceous monocots: a phylogenetic evaluation. *Sbornik Narodniho Muzea v Praze Rada B Prirodni Vedy* **64**, 59-87 (2008).
88. Doyle, J. A. & Robbins, E. I. Angiosperm pollen zonation of the continental Cretaceous of the Atlantic coastal plain and its application to deep wells in the Salisbury Embayment. *Palynology*, 43-78 (1977).
89. Doyle, J. A. Recognising angiosperm clades in the Early Cretaceous fossil record. *Historical Biology* **27**, 414-429 (2015).
90. Doyle, J. A. & Endress, P. K. Integrating Early Cretaceous Fossils into the Phylogeny of Living Angiosperms: ANITA Lines and Relatives of Chloranthaceae. *International Journal of Plant Sciences* **175** (2014).
91. Mohr, B. A. R. & Bernardes-de-Oliveira, M. E. C. Endressinia brasiliana, a magnolialean angiosperm from the lower Cretaceous Crato Formation (Brazil). *International Journal of Plant Sciences* **165**, 1121-1133 (2004).
92. Mohr, B. A. R., Coiffard, C. & Bernardes-de-Oliveira, M. E. C. Schenkeriphyllum glanduliferum, a new magnolialean angiosperm from the Early Cretaceous of Northern Gondwana and its relationships to fossil and modern Magnoliales. *Review of Palaeobotany and Palynology* **189**, 57-72 (2013).
93. Sauquet, H. *et al.* Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* **142**, 125-186 (2003).
94. Smith, S. Y. & Stockey, R. A. Establishing a fossil record for the perianthless Piperales: *Saururus tuckerae* sp. nov. (Saururaaceae) from the Middle Eocene Princeton Chert. *American Journal of Botany* **94**, 1642–1657 (2007).
95. Rouse, G. E. & Mathews, W. H. Radioactive dating of Tertiary plant-bearing deposits. *Science* **133**, 1079-1080 (1961).
96. Mathews, W. H. Potassium-argon age determinations of Cenozoic volcanic rocks from British Columbia. *Geological Society of America Bulletin* **75**, 465-468 (1964).
97. Hills, L. V. & Baadsgaard, H. Potassium-argon dating of some Lower Tertiary strata in British Columbia. *Bulletin of Canadian Petroleum Geology* **15**, 138-149 (1967).
98. Read, P. B. Geology and industrial minerals of the Tertiary basins, south-central British Columbia. *British Columbia Geological Survey Geo-File* **2000-3** (2000).

99. Moss, P. T., Greenwood, D. R. & Archibald, S. B. Regional and local vegetation community dynamics of the Eocene Okanagan Highlands (British Columbia – Washington State) from palynology. *Canadian Journal of Earth Sciences* **42**, 187-204 (2005).
100. Hochuli, P. A., Heimhofer, U. & Weissert, H. Timing of early angiosperm radiation: recalibrating the classical succession. *Journal of the Geological Society* **163**, 587-594 (2006).
101. Hughes, N. F. *The enigma of angiosperm origins*. (Cambridge University Press, 1994).
102. Berry, E. W. The Upper Cretaceous and Eocene floras of South Carolina and Georgia. *United States Geological Survey Professional Paper* **84**, 1-200 (1914).
103. Hertweck, K. L. *et al.* Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society* **178**, 375-393 (2015).
104. Iles, W. J. D., Smith, S. Y., Gandolfo, M. A. & Graham, S. W. Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* **178**, 346-374 (2015).
105. Campbell, B. G. & Gohn, G. S. Stratigraphic framework for geologic and geohydrologic studies of the subsurface Cretaceous section near Charleston, South Carolina. *United States Geological Survey Map MF-2273*, 1-11 (1994).
106. Habib, D. & Miller, J. A. Dinoflagellate species and organic facies evidence of marine transgression and regression in the atlantic coastal plain. *Palaeogeography, Palaeoclimatology, Palaeoecology* **74**, 23-47 (1989).
107. Christopher, R. A. & Prowell, D. C. A palynological biozonation for the uppermost Santonian and Campanian Stages (Upper Cretaceous) of South Carolina, USA. *Cretaceous Research* **31**, 101-129 (2010).
108. Friis, E. M. *Spirematospermum chandlerae* sp. nov., an extinct species of Zingiberacea from the North American Cretaceous. *Tertiary Research* **9**, 7-12 (1988).
109. Gandolfo, M. A., Nixon, K. C. & Crepet, W. L. Triuridaceae fossil flowers from the Upper Cretaceous of New Jersey. *American Journal of Botany* **89**, 1940-1957 (2002).
110. Courtillot, V. E. & Renne, P. R. On the ages of flood basalt events. Comptes Rendus Geoscience 335, 113–140 (2003).
111. Dilcher, D. L., Sun, G., Ji, Q. & Li, H. Q. An early infructescence Hyrcantha decussata (comb. nov.) from the Yixian Formation in northeastern China. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9370-9374 (2007)
112. Wang, W., Dilcher, D. L., Sun, G., Wang, H.-S. & Chen, Z.-D. Accelerated evolution of early angiosperms: Evidence from ranunculalean phylogeny by integrating living and fossil data. *Journal of Systematics and Evolution* **54**, 336-341, (2016).
113. Crepet, W. L., Nixon, K. C. & Daghlian, C. P. Fossil Ericales from the Upper Cretceous of New Jersey. *International Journal of Plant Sciences* **174**, 572-584 (2012).
114. Barreda, V. D., Palazzesi, L., Tellería, M. C., Olivero, E. B., Raine, J. I. & Forest, F. Early evolution in the angiosperm clade Asteraceae in the Cretaceous of Antarctica. *Proceedings of the National Academy of Sciences USA* **112,** 10989-10994 (2015).
115. Panero, J. L. Phylogenetic uncertainty and fossil calibration of Asteraceae chronograms. *Proceedings of the National Academy of Sciences USA* **113,** E411 (2016).
116. Martínez-Millán, M. Fossil record and age of the Asteridae. *Botanical Review* **76**, 83-135 (2010).
117. Zavada, M. & de Villiers, S. Pollen of the Asteraceae from the Paleocene-Eocene of South Africa. *Grana* **39**, 39-45 (2000).

118. Wade, B. S., Pearson, P. N., Berggren, W. A. & Pälike, H. Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the geomagnetic polarity and astronomical time scale. *Earth-Science Reviews* **104**, 111-142 (2011).
119. Takahashi, M., Crane, P. R. & Ando, H. *Esgueiria futabensis* sp. nov., a new angiosperm flower from the Upper Cretaceous (lower Coniacian) of northeastern Honshu, Japan. *Paleontological Research* **3**, 81-87 (1999).
120. Friis, E. M., Pedersen, K. R. & Crane, P. R. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeography, Palaeoclimatology, Palaeoecology* **232**, 251-293 (2006).
121. Yazykova, E. Ammonite and inoceramid radiations after the Santonian–Campanian bioevent in Sakhalin, Far East Russia. *Lethaia* **35**, 51-60 (2002).
122. McClain, A. M. & Manchester, S. R. Dipteronia (Sapindaceae) from the Tertiary of North America and Implications for the Phytogeographic History of the Aceroideae. *American Journal of Botany* **88**, 1316-1325 (2001).
123. Qiu, Y. L. *et al.* A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* **168**, 691-708 (2007).
124. Manchester, S. R., Chen, Z.-D., Lu, A.-M. & Uemura, K. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *Journal of Systematics and Evolution* **47**, 1-42, (2009).
125. Nichols, D. J. & Ott, H. L. Biostratigraphy and evolution of the *Momipites-Caryapollenites* lineage in the early Tertiary in the Wind River Basin, Wyoming. *Palynology* **2**, 93-112 (1978).
126. Peppe, D. J. Megafloral change in the early and middle Paleocene in the Williston Basin, North Dakota, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology* **298**, 224-234, (2010).
127. Vandenberghe, N., Hilgen, F. J. & Speijer, R. P. in *The geologic timescale 2012* (eds F. M. Gradstein, J. G. Ogg, M. Schmitz, & G. Ogg) 855-921 (Elsevier, 2012).
128. Boucher, L. D., Manchester, S. R. & Judd, W. S. An extinct genus of Salicaceae based on twigs with attached flowers, fruits, and foliage from the Eocene Green River Formation of Utah and Colorado, USA. *American Journal of Botany* **90**, 1389-1399 (2003).
129. Manchester, S. R., Judd, W. S. & Handley, B. Foliage and fruits of early poplars (Salicaceae: Populus) from the Eocene of Utah, Colorado, and Wyoming. *International Journal of Plant Sciences* **167**, 897-908, (2006).
130. Smith, M. E., Carroll, A. R. & Singer, B. S. Synoptic revision of a major ancient lake system: Eocene Green River Formation, western United States. *GSA Bulletin* **120**, 54-84, (2008).

Appendix 3.

A description of characters, their states and rationale for scoring. Characters derived from published studies are indicated in parentheses by an author abbreviation, followed by the original character number. Abbreviated authors are as follows: D = ; DE = Doyle and Endress (2010); G = Gensel (1992); GRD = Garbary *et al.* (1993); HB = Hilton and Bateman (2006); HX = Hao and Xue (2013); KC = Kenrick and Crane (1997); M = Mishler *et al.* (1994); N = Newton *et al.* (2000) ; R = Rothwell (1999); RG = Renzaglia and Garbary (2001); S = Schneider *et al.* (2009); X = Xue *et al.* (2010). Modified characters are indicated by 'MOD' following the character number and underlined text. Combined characters are indicated by 'COMB' following the character number and underlined text. New characters are indicated by 'NEW' following the character number. Unless otherwise stated, characters are scored as the original reference.

I. General characteristics (1 - 18)

1. (M1) **Habitat of free-living vegetative stage: Freshwater (0); Brackish or marine (1); Terrestrial (2).**
Green algae are mostly either living in freshwater or brackish to marine environments (Mishler *et al.* 1994). Some green algae are polymorphic, occurring as freshwater and terrestrial (soil or epiphytic) forms. Most extant embryophytes are assumed to be terrestrial, with the exceptions of water ferns (*Azolla, Marsilea*), Nymphaeaceae (Water lilies), *Ceratophyllum*, *Pistia*, *Aponogeton* and *Nelumbo*. Extinct embryophyte taxa with insufficient information are scored unknown, particularly the early flowering plants. There is some speculation whether *Taeniocrada dubia* was in fact aquatic due to flattened axes, dense mats, apparent lack of stomata and the rock matrix in which they are deposited (Taylor *et al.* 2009). The same is true of *Catenalis digitata* (Hao and Beck 1991), but there is not enough definitive evidence.

2. (M5) **Growth form: Unicellular or coccoid (0); Multicellular (1); Coenobic (2).**
All living and extinct embryophytes are multicellular. Green algae may be unicellular, multicellular or occur as a colony (Mishler *et al.* 1994).

3. MOD. (M2) **Life history: Haplontic (0); Diplontic (1); <u>Isomorphic haplodiplontic</u> (2); <u>Heteromorphic haplodiplontic</u> (3).**
Niklas and Kutschera (2010) consider the life cycle of unicellular green algae to be fundamentally the same as those with multicellular phases, defined either by zygotic or gametic meiosis. However, in all unicellular taxa selected, their life cycles are currently unknown, although *N. olivacea* is known to exhibit zygotic meiosis (Suda and Watanabe 2004). Most members of Chlorophyceae have a haplontic life cycle. Some members of Ulvophyceae have a haplodiplontic life cycle. *Ulothrix* is unusual because of an additional unicellular *Codiolum* stage (diploid zygote). The shift from an isomorphic to heteromorphic haplodiplontic life cycle was a key innovation for embryophytes. This is thought to have occurred independently from some members of the Ulvophyceae (Niklas and Kutschera 2010). All living embryophytes are scored as having heteromorphic diplohaplontic life cycles. For many of the basal extinct taxa, the gametophyte, and thus life history, is unknown. The gametophytes are known for some of the Rhynie Chert plants, and indicate a more or less isomorphic haplodiplontic life cycle i.e. axial gametophytes (*Horneophyton, Aglaophyton, Stockmansella, Rhynia* and *Nothia*).

4. COMB (HB1, KC5.20, S81) **Sporogenesis: Homosporous (0); Heterosporous (1).**
Sporogenesis is a developmental process unique to the embryophytes by which haploid cells resulting from meiosis are covered with a sporopollenin-impregnated wall (a sporoderm) (Brown and Lemmon 2011b), thus this character is inapplicable to all green algae. Homosporous plants produce one type of spore, including all extant bryophytes,

members of Lycopodiales and the majority of ferns. Heterosporous plants produce two types of spores, differentiated by sex, including all extant gymnosperms and angiosperms, *Isoetes, Selaginella* and members of Salviniales. Extinct taxa are scored according to Kenrick and Crane (1997) and Hilton and Bateman (2006).

5. (S131) **Gametophyte/sporophyte life-span: Gametophyte phase dominant, gametophyte long-lived, sporophyte short-lived (0); Gametophyte and sporophyte both long-lived (1); Sporophyte phase dominant, sporophyte long-lived, gametophyte short-lived (2).**
Rule 1*. Contingent on alternation of generations (a haplodiplontic life cycle; character 3, states 2, 3).*
This character is inapplicable to all green algae, with the exception of *Ulothrix*. The gametophyte phase is dominant in all bryophytes, while the sporophyte is dominant in the tracheophytes. The relative duration of each phase varies between lineages. The gametophyte phase is short-lived in seed plants and most ferns, and long-lived in Equisetaceae, Lycopodiaceae, Ophioglossaceae and Psilotaceae and some early fern groups (Schneider *et al.* 2009). It is not possible to determine this character for extinct taxa, thus are scored as unknown.

7. (M3) **Vegetative cell or thallus attached to substrate: No (0); Yes (1).**
All embryophytes and some green algae, including *Chara* and *Coleochaete*, attach to the substrate. Some zygnematales, although predominantly free-living, occasionally attach to substrates via rhizoidal processes (e.g. *Spirogyra, Zygnema*).

8. (M4) **Multicellular radial symmetry: No (0); Yes (1).**
Rule 2*. Contingent on a multicellular or coenobic growth form (character 2, states 1, 2).*
This character refers only to the radial symmetry observed in some Dasycladales algae. All living and extinct embryophytes are scored (0). Unicellular algae are scored as inapplicable.

9. (M76) **Apical cell growth: No (0); Yes (1).**
Rule 3*. Contingent on a multicellular or coenobic growth form (character 2, states 1, 2).*
Unicellular algae are scored inapplicable. All living and extinct embryophytes are scored (1), due to the presence of apical cell growth in structures such as rhizoids, pollen tubes, root hairs and trichomes. Some epiphytic algae (*Trentopohlia*, *Cephaleuros*), and some charophytes (*Chara*, *Coleochaete*) also exhibit apical /terminal cell growth. There is a suggestion of tip growth in *Chaetosphaeridium*, but Thompson (1969) did not consider this form of growth to be equivalent to the terminal meristematic growth exhibited in *Coleochaete*. Apical cell growth occurs only in the gametophytes of liverworts and hornworts and during early embryogenesis in moss sporophytes. It is lost in the gametophyte generation of seed plants and members of Selaginellales and Isoetales.

10. NEW Number of cutting faces of apical cells: Less than 3 (0); Three or more (1).
Rule 4*. Contingent on apical cell growth (character 9, state 1).*
In all embryophytes apical cells possess 3 or more cutting faces, producing three-dimensional tissues. Charophytes with apical/ terminal meristems (*Chara* and *Coleochaete*) have less than 3 cutting faces (Graham and Wilcox 2000).

11. (M6) **Vegetative cells contiguous in multicellular organism: No (0); Yes (1).**
Rule 5*. Contingent on a multicellular or coenobic growth form (character 2, states 1, 2).*
Unicellular algae are inapplicable. The lack of contiguous vegetative cells only applies to some Chlamydomonadales algae. All living and extinct embryophytes are scored (1).

12. (M7) **Multinucleate vegetative cells: No** (**0**)**; Yes** (**1**)**.**
This character refers to the multinucleate cells known in some green algae, and not in embryophytes. All extant embryophytes are scored (0). Extinct taxa are scored unknown.

13. (M8) **Coenocytic: No** (**0**)**; Yes** (**1**)**.**
Rule 6*. Contingent on multinucleate vegetative cells (character 12, state 1).*
Multinucleate coenocytes occur when nuclear division is not followed by cytokinesis. This character applies to the organism as a whole and so embryophytes are scored (0). Extinct taxa are scored unknown.

14. (M9) **Distromatic foliar thalli: Absent** (**0**)**; Present** (**1**)**.**
Rule 7*. Contingent on a multicellular or coenobic growth form (character 2, states 1, 2).*
This character refers to thalli of certain green algal taxa. All living and extinct embryophytes are scored (0). Unicellular algae scored inapplicable.

15. (M10) **Plasmodesmata: Absent** (**0**)**; Present** (**1**)**.**
Rule 8*. Contingent on a multicellular or coenobic growth form (character 2, states 1, 2).*
Plasmodesmata occur between the cells of all living embryophytes and some green algae (Trentepohliales, Chaetophorales, *Chara*, *Chaetosphaeridium*, *Coleochaete*). Pickett-Heaps (1975) notes that plasmodesmata only form in cross walls that use cell plates during mitosis. All living embryophytes are scored (1), while extinct taxa are scored unknown.

16. NEW Desmotubules within plasmodesmata: Absent (**0**)**; Present** (**1**)**.**
Rule 9*. Contingent on the presence of plasmodesmata (character 15, state 1).*
Internal tubular endoplasmic reticulum (desmotubules) observed within the plasmodesmata of *Chara* are homologous to those of higher plants, after Cook *et al.* (1997). Not present in the Trentepohliales Chapman and Good (1978) or *Coleochaete* (Marchant and Pickett-Heaps 1973). In ultrastructure studies of a few bryophytes, desmotubules are reported in *Monoclea, Notothylas* and *Sphagnum* (Cook *et al.* 1997) and *Plagiomnium* and *Atrichum* (Ligrone and Duckett 1994). All bryophytes and tracheophytes are scored (1).

17. (M11) **Parenchyma: No** (**0**)**; Yes** (**1**)**.**
Rule 10*. Contingent on a multicellular or coenobic growth form (character 2, states 1, 2).*
As the ground tissue of plants, all living and extinct embryophytes are scored (1). A parenchymatous tissue is also observed in *Coleochaete* and is deemed homologous to that of the land plants (Graham 1982).

18. (M12) **Vegetative cells form filaments: No** (**0**)**; Yes, unbranched** (**1**)**; Yes, branched** (**2**)**; Yes, multi-axial** (**3**)**.**
All living and extinct embryophytes are scored (0) as this character is primarily a means of distinguishing filamentous algae.

II. Molecular and Mitotic (**19 - 42**)

19 – 20. ***Cell walls and extracellular matrix***

19. (M52) **Common matrix surrounding cells: Absent** (**0**)**; Present** (**1**)**.**
Rule 11*. Contingent on coenobic growth form (character 2, state 2).*

This character only applies to certain colonial algae in which cells that share an extracellular gelatinous matrix. Extant and extinct embryophytes are scored inapplicable.

20. (M54) **Crystalline cell wall(s): Absent** (**0**)**; Present** (**1**)**.**
This refers to the hydroxyproline-rich glycoprotein cell walls that are characteristic of certain groups of chlorophycaean algae (Chlamydomonadales), but not in charophytes or embryophytes, which are scored as absent (0) (Domozych *et al.* 2012; Domozych and Domozych 2014). Extinct taxa are scored unknown.

21 – 29. ***Organelles***

21. NEW Chloroplast number per cell: None (**0**)**; One** (**1**)**; Multiple** (**2**)**.**
Tracheophytes, mosses and liverworts all possess multiple chloroplasts per cell. Hornwort cells posses 1 chloroplast (Renzaglia *et al.* 2009), occasionally more than 1 (*Megaceros*); variable in green algae. The cells of *Prototheca* lack chloroplasts. Extinct taxa are scored unknown.

24. NEW Multiple disc-shaped chloroplasts: Absent (**0**)**; Present** (**1**)**.**
All embryophytes were scored as possessing chloroplasts consisting of multiple discs (Bell and Hemsley 2000). Most green algae do not possess multiple discs, with the exception of *Chara* and certain Ulvophyceans. Extinct taxa are scored unknown.

25. NEW Thylakoid stacking apparatus: Extensive parallel lamellae bands (**0**)**; Irregularly stacked thylakoids** (**1**)**; Regularly stacked thylakoids** (**grana**)**, without end membranes** (**2**)**; Regularly stacked thylakoids** (**grana**)**, with end membranes** (**3**)**.**
Thylakoids in the chloroplasts of green algae may be irregularly stacked or occur in bands or laminar structures. Regularly stacked thylakoids in grana are characteristic of embryophytes. Hornwort thylakoid systems are regularly stacked in channels and lack end membranes (Renzaglia *et al.* 2000), which is also the case in some charophytes (*Chara, Coleochaete*) (Vaughn *et al.* 1992; Graham and Wilcox 2000). Extinct taxa are scored unknown.

26. (M22) **Pyrenoids: Absent** (**0**)**; Present** (**1**)**.**
Pyrenoids are defined here as Rubisco-rich bodies within the chloroplasts of some green algae and some hornworts (Villarreal and Renner 2012). They are present in all charophytes with the exception of *Chara* (Bell and Hemsley 2000). They are not present in the chloroplasts of liverworts, mosses or tracheophytes. Extinct taxa are scored unknown.

27. NEW Number of pyrenoids: One (**0**)**; multiple** (**1**)**.**
Rule 12*. Contingent on the presence of pyrenoids (character 26, state 1).*
Those taxa lacking pyrenoids are scored inapplicable. Where present, hornworts only possess one pyrenoid (Villarreal and Renner 2012). Green algae can possess 1 or multiple pyrenoids (Graham and Wilcox 2000). Extinct taxa are scored unknown.

28. (M23) **Thylakoid membrane transverse pyrenoid: Absent** (**0**)**; Present** (**1**)**.**
Rule 13*. Contingent on the presence of pyrenoids (character 26, state 1).*
Those taxa lacking pyrenoids are scored inapplicable. Pyrenoids within the chloroplasts of certain hornworts are traversed by thylakoids, producing a 'multiple pyrenoid formation' (Renzaglia *et al.* 2009; Villarreal and Renner 2012). In some green algae, thylakoids are known to penetrate the pyrenoid, and in some cases traverse, producing subunits, but are not as prolific as in hornworts, thus scored absent. Extinct taxa are scored unknown.

29. (M56) **Contractile vacuoles: Two** (**0**)**; More than two** (**1**)**; One** (**2**)**; Absent** (**3**)**.**
Contractile vacuoles are organelles involved in hydrostatic equilibrium in unicellular algae and zoospores that occur in some algae (e.g. *Chlamydomonas*). All embryophytes are scored absent. Extinct taxa are scored unknown.

30 – 42. ***Cell division***

30. (M35) **Closed mitotic spindle: Absent** (**0**)**; Present** (**1**)**.**
In most green algae, the nuclear envelope persists throughout mitosis i.e. closed mitosis (with some exceptions in Prasinophyaceae). All embryophytes and charophytes are scored absent following Hodson and Bryant (2012): "In both charophytes and embryophytes, the mitotic spindle is persistent and mitosis is open". Extinct taxa are scored unknown.

31. (M36) **Spindles collapse during telophase: Absent** (**0**)**; Present** (**1**)**.**
The spindle collapses at telophase in members of Chlorodendrales (*Tetraselmis*) (Lewis and McCourt 2004), Trebouxiophyceae and Chlorophyceae. Spindles are persistent in Pedinophyceae, Ulvophyceae, and all embryophytes and charophytes (Hodson and Bryant 2012). Extinct taxa are scored unknown.

32. NEW Centrioles in vegetative mitosis: Absent (**0**)**; Present** (**1**)**.**
Centrioles are organelles involved in cell division, mostly in animal cells. In the plant kingdom, centrioles are only present in vegetative mitosis of certain chlorophytes and coleochaetales (Brown and Lemmon 2011a) . All other charophytes and embryophytes are scored absent (0). This does not include the centrioles involved in the division of male gametes of certain charophytes, bryophytes, and embryophytes. Extinct taxa are scored unknown.

33. MOD (M37) **<u>Metacentric spindle:</u>** (**cupping microtubules surround the centrioles during mitosis**)**: Absent** (**0**)**; Present** (**1**)**.**
Rule 14*. Contingent on the presence of centrioles (character 32, state 1).*
Those taxa without centrioles are scored inapplicable. A metacentric spindle is only present in Trebouxiophyceae and some Chlorodendrales (Graham and Wilcox 2000). All embryophytes are scored absent. Extinct taxa are scored as unknown.

35. NEW Acentriolar MTOCs in vegetative mitosis: Polar organisers (**PO**) (**0**)**; axial microtubular system associated with plastid** (**AMS**) (**1**)**; Associated with the nuclear-envelope** (**NE**) (**2**)**.**
Rule 15*. Contingent on the lack of centrioles (character 32, state 0).*
Those taxa with centrioles are scored inapplicable. In the absence of centrioles in embryophyte cells, microtubules are organised and spindles initiated from 'microtubules organising centres' (MTOCs) (Shimamura *et al.* 2004; Brown and Lemmon 2011a). In liverworts, microtubules initiate from polar organisers (POs). In hornworts, an axial microtubule system is associated with the division of the single plastid. In mosses and all other embryophytes, the MTOCs are associated with the nuclear envelope. In *Closterium*, mitosis and chloroplast division is by constriction mediated by actin microfilaments (Hashimoto 1992). Extinct taxa are scored unknown.

36. NEW Vegetative cell division by cleavage furrow: Absent (**0**)**; Present** (**1**)**.**
Cytokinesis of vegetative cells in many green algae is achieved through cleavage furrowing, which may or may not be accompanied by a phycoplast (see character 37). All embryophytes are scored absent. Extinct taxa are scored unknown.

37. MOD (M38) **<u>Phycoplast:</u>** (**microtubule formation in the plane of cell division**)**: Absent** (**0**)**; Present** (**1**)**.**

The alignment of the microtubules in the plane of cell division is characteristic of a phycoplast, present only in some chlorophyte algae. All charophytes and embryophytes are scored absent. Extinct taxa are scored unknown.

38. MOD (M39) **Phragmoplast: Absent** (**0**)**; Present** (**1**)**; <u>Present, but controlled by cleavage</u>** (**<u>2</u>**)**.**
Cell division in embryophytes, Charales and Coleochaetales is characterised by the presence of a phragmoplast and a cell plate (Doyle 2013). This contrasts with modes of cell division in the chlorophytes, which involves either cleavage furrow and/or a phycoplast (characters 36 and 37). Exceptions are the Trentepohiales (*Trentepohlia, Cephaleuros*) which possess a phragmoplast (Chapman and Borkhsenious 2001). The Zygnemataceae (*Spirogyra* etc.) also possess a phragmoplast, but it is more primitive than those observed in land plants (Fowke and Pickett-Heaps 1969; Pickett-Heaps and Wetherbee 1987; Pickett-Heaps *et al.* 1999) and is functionally distinct from those of higher plants (Sawitzky and Grolig 1995). Pickett-Heaps *et al.* (1999) describe this type of division as '...cytokinesis initiated by an ingrowing furrow lined with actin. Once this furrow impinges on the persistent overlapping spindle MTs, a small phragmoplast develops at the region of contact.' Extinct taxa are scored unknown.

39. (M40) **Cell plate in cytokinesis: Absent** (**0**)**; Present** (**1**)**.**
A cell plate is present in all embryophytes during cytokinesis (Doyle 2013). The cell plate is absent in most green algae, except for those with a phragmoplast. Extinct taxa are scored unknown.

40. NEW Monoplastidic mitosis: Present (**0**)**; Absent** (**1**)**; Spermatogenesis only** (**2**)**.**
Monoplastidic mitosis occurs in green algae, including *Coleochaete* and *Chara*. Mitosis is monoplastidic in hornworts and the moss *Takakia*; in all other mosses vegetative mitosis is polyplastidic, although monoplastidic in the final division of spermatogenesis (Brown and Lemmon 1990; Renzaglia *et al.* 2009). In most liverworts mitosis is polyplastidic, except for *Monoclea*, *Blasia* (Brown and Lemmon 1992). Mitosis in most tracheophytes is polyplastidic, with exception of *Isoetes* and *Selaginella*. Extinct taxa are scored unknown.

41. MOD (S83) **Monoplastidic meiosis: Present** (**0**)**; Absent** (**1**)**; <u>Spermatogenesis only</u>** (**<u>2</u>**)**.**
Monoplastidic meiosis occurs in green algae, including *Coleochaete and Chara*. Meiosis is monoplastidic in hornworts and mosses, including *Megaceros*, which possesses multiple plastids in mature cells, but division of the chloroplast is monoplastidic during meiosis (Brown and Lemmon 1990; Vaughn *et al.* 1992). Conversely, in most liverworts meiosis is polyplastidic, with the exception of *Monoclea*, *Blasia* and some species of *Haplomitrium* and during spermatogenesis in *Lunularia* and *Marchantia* (Renzaglia *et al.* 1994; Shimamura *et al.* 2003; Brown and Lemmon 2013). Meiosis in most tracheophytes is polyplastidic, with the exception of *Isoetes*, *Selaginella*. Monoplastidic meiosis is seen during spermatogenesis of *Lycopodium* and certain eusporangiate ferns. Extinct taxa are scored unknown.

42. (M51) **Sporulation vs. Zellteilung: Sporulation** (**0**)**; Zellteilung** (**1**)**.**
Two types of cell division occur in green algae: sporulation (cytogony), in which the daughter cells are motile, or with evidence of derivation from flagellated condition; or 'Zellteilung' (cytogamy), in which the daughter cells originate from vegetative cells (Sluiman *et al.* 1989). All embryophytes are coded (1). Extinct taxa are scored unknown.

III. Sporophyte (**43 - 299**)

43 – 51. ***General sporophyte features and development***

43. (KC3.1) **Multicellular sporophyte: Absent** (**0**)**; Present** (**1**)**.**
A multicellular sporophyte (diploid) generation is characteristic of all embryophytes and does not occur in charophytes or chlorophytes, with the exception of certain members of the ulvophyceans.

44. COMB (KC 3.2, KC4.1, M100, S23, HX1) **Independent multicellular sporophyte: Dependent** (**0**)**; Independent** (**1**)**.**
Rule 16*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
Most algae lack a multicellular diploid generation, due to their diplontic or haplontic life cycles, hence scored inapplicable. In mosses, liverworts and hornworts the sporophyte remains in intimate contact with the gametophyte. In all other extant lineages, it is independent. Unequivocal evidence for sporophyte independence in many fossil polysporangiophytes, particularly the minute taxa, has not been demonstrated. Extinct taxa with evidence of a clearly defined gametophyte (e.g. *Horneophyton ligieri*) or belong to a group known to possess rhizomatous axes (e.g. zosterophylls) are scored as having an independent gametophyte. Otherwise protracheophytes, some basal tracheophytes and cryptophytes are scored unknown. The sporophyte of *Cooksonia parenensis* is considered to be independent, after Gerrienne *et al.* (2006).

45. MOD (**KC3.5**) **Sporophyte growth via an intercalary meristem: Absent** (**0**)**; Present** (**1**)**.**
Rule 17*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
This method of growth is a uniquely derived feature of hornworts (Villarreal and Renzaglia 2015).

46. MOD (KC 3.3, KC4.2, HX2) **Persistent sub-sporangial stalk formed by cell division: Absent** (**0**)**; Present** (**1**)**.**
Rule 18*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
Modified after "Well-developed sporangiophore', defined by Kenrick and Crane (1997) as a persistent sporophyte axis with internal differentiation of tissues, thus considered the unbranched setae of mosses to be homologous to sporangium-bearing axes of polysporangiates (Kenrick and Crane 1997; Tomescu *et al.* 2014). Here we also consider the setae of mosses and the axes of tracheophytes to be homologous but based on growth and development via cell division and not based on function. The setae of liverworts are scored absent because of the ephemeral nature and differences in development of sporangial stalks. There is no such homologous structure in hornworts, which develops through cell expansion (see character 45).

48. NEW Determinate vs. indeterminate meristematic growth: Determinate (**0**)**; Indeterminate via apical meristem** (**1**)**; Indeterminate via basal meristem** (**2**)**.**
Rule 19*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
The growth of bryophyte sporophytes is determinate, i.e. via a basal or intercalary meristem. Development of the shoot apical meristem in tracheophytes allowed for their indeterminate growth. The growth of earliest polysporangiophytes is unknown. Apical meristematic cells are evidenced in *Rhynia* (Kidston and Lang 1917; Edwards 2004) and *Asteroxylon* (Hueber 1992).

49. MOD (S50) **Shoot apical meristem structure in developing sporophyte axis: Unicellular or up to 4 initial cells** (**0**)**; Complex and more than 4 initial cells** (**1**)**.**
Rule 20*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
Although *Trentepohlia* and *Cephaleuros* do exhibit apical growth via an apical cap, they do not possess apical meristems equivalent to higher plants, thus scored inapplicable. All bryophytes possess unicellular apical meristems in developing sporophytes (Goffinet and Buck 2013). Ferns, with the exception of Marratiales and Osmundales, possess simple unicellular meristems (Schneider *et al.* 2009). In the latter, the single cell may be replaced with up to 4 initials. Lycophytes possess complex meristems lacking a clear initial, with the exception of *Selaginella* (Schneider *et al.* 2009).

Seed plants possess complex, multi-structured meristems. All extinct taxa are scored unknown with the exception of *Rhynia* (Kidston and Lang 1920; Edwards 1994) and *Asteroxylon* (Hueber 1992) in which complex apical meristems have been illustrated.

50. MOD (KC3.6, S129, N33) **Foot shape: Cup-shaped (0); Bulbous (1); Short conical (2); Long conical / tapering (3).**
Rule 21*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
Foot shape is highly variable among the bryophytes. *Porella pinnata* was described as possessing a club-shaped (bulbous) foot (Manning 1914), as with *Marchantia*, whereas *Conocephalum conicum* has a conical (tapered) foot (Graham, 1909). The Ricciaceae lack a foot (Shaw and Renzaglia 2004). Hornworts possess generally globose feet (Goffinet and Shaw 2009). All extant lycophytes and ferns have a bulbous foot. Taxa without a foot (seed plants and green algae) are scored inapplicable. Extinct taxa are scored unknown.

51. (HB16) **Apical meristem tunica: Absent (0); Present (1).**
Rule 22*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
An apical meristem tunica is only known in angiosperms, thus all other applicable taxa are scored absent, with the exception of the Araucariaceae (Hilton and Bateman 2006). All angiosperms, with the exception of *Welwitschia*, are scored present, although this may be controversial as it has been poorly surveyed among angiosperms (Friis *et al.* 2011). The presence of an apical meristem tunica amongst basal angiosperms has been confirmed by Posluszny and Tomlinson (2003) and (Doyle 2006). Extinct taxa are scored unknown.

52 – 60. ***Epidermal cells and stomata***

52. COMB (S29, M105/107, R23) **Shoot indumentum / trichomes: Absent (0); Present (1).**
Rule 23*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
Absent in all bryophyte sporophytes and green algae. Almost all extant tracheophyte sporophytes produce unicellular or multicellular epidermal outgrowths / trichomes in the form of hairs and/or scales and other structures that may be glandular. Indument may be ephemeral in some taxa. The scoring of ferns is after (Schneider *et al.* 2009). Where preservation allows good cellular or cuticular preservation, extinct taxa are scored accordingly. Extinct taxa lacking clear structures are scored unknown.

53. MOD (S30) **Scales: Absent (0); Present (1).**
Rule 24*. Contingent on the presence of shoot indumentum (character 52, state 1).*
All taxa without shoot indumentum are scored inapplicable. Scales are defined as modified hairs known to occur on some ferns. This character is modified from the original, as there was uncertainty about the presence of hairs on some taxa.

54. (S58) **Mucilage-producing hairs: Absent (0); Present (1).**
Rule 25*. Contingent on the presence of shoot indumentum (character 52, state 1).*
All taxa without shoot indumentum are scored inapplicable. Mucilage-producing hairs are a character of certain groups of ferns (Schneider *et al.* 2009) and a feature of Cabombaceae (Endress 2005). Extinct taxa are scored unknown.

55. MOD (KC3.14, M83)**-Extra-mural material: Absent (0); Present - differentiated (1); Present - undifferentiated (2).**
Extra-mural material includes cuticle and cuticle-like coverings. Green algae lack any extra-mural coverings. *Coleochaete* possesses a surface layer with similar construction, but is not considered to be a true cuticle. Some

liverworts possess a thin undifferentiated layer of extra mural material (EMM), while hornwort and moss sporophytes have a more complex, differentiated EMM (Ligrone *et al.* 2012). All tracheophytes possess a cuticle, so scored (1). Extinct taxa without explicit description of a cuticle or extra mural material were scored unknown.

56. NEW Schizogenous intercellular spaces: Absent (0); Present (1).
Rule 26*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
The spaces between two or more cells in tissue, specifically those originating from the separation of cell walls along a middle lamella i.e. does not include spaces as a result of breakdown or tearing. These are essential for aeration of internal tissues. These spaces are not present in green algae or the diploid phase of most liverworts. They are present in the diploid phase of all tracheophytes and the majority of mosses and hornworts (Raven 1996; Renzaglia *et al.* 2000). Extinct taxa without described intercellular spaces are scored unknown.

57. NEW Content of schizogenous intercellular spaces: Initially liquid (0); Gas (1); Always liquid (2).
Rule 27*. Contingent on the presence of schizogenous intercellular spaces (character 56, state 1).*
Those taxa without schizogenous intercellular spaces are scored inapplicable. In tracheophytes they are air-filled throughout stomatal ontogeny. In liverworts that possess spaces, they are always filled with liquid. In hornworts and mosses with schizogenous intercellular spaces they are initially mucilage-filled while stomata are closed, drying out with increasing sporophyte maturity (Pressel *et al.* 2014). Extinct taxa are scored unknown.

58. COMB (KC3.15, KC4.12, HB11, HB12, M89) **Stomata: Absent (0); Present (1).**
True stomata are defined as openings in the epidermis, each bordered by two guard cells with the function of gas exchange. The presence of stomata was assumed for all extant higher land plants. Liverwort sporophytes do not possess stomata. Stomata are a plesiomorphic character of hornwort sporophytes, but are lost in *Notothylas*, *Megaceros*, *Nothoceros*, *Dendroceros* (Renzaglia *et al.* 2009). Stomata are present in some moss sporophytes; absent in *Takakia*, *Sphagnum, Andraea*, *Andreaeobryum*, *Atrichum, Tetraphis, Scouleria, Leucobryum* and *Fontinalis* (Goffinet *et al.* 2009). Pseudo-stomata (stomata-like structures, including guard cells) present in *Sphagnum* are not considered to be true stomata, thus scored absent. Extinct taxa lacking explicit description are scored unknown.

59. COMB (DE36, HB12) **Stomata type: Paracytic (0); Laterocytic (1); Anomocytic (2); Stephanocytic (3).**
Rule 28*. Contingent on the presence of stomata (character 58, state 1).*
All taxa without stomata are scored inapplicable. All stomatiferous hornworts and mosses possess anomocytic stomata, which also occur in many of the early extinct taxa. Ferns are scored according to (Schneider *et al.* 2009) and seed plants according to (Hilton and Bateman 2006). Additional information was provided by Rudall *et al.* (2013).

60. (S57) **Origin of subsidiary cells: Perigenous (0); Mesogenous (1); Mesoperigenous (2).**
Rule 29*. Contingent on the presence of stomata (character 58, state 1).*
All taxa without stomata are scored inapplicable. Some scoring was changed from the original according to Rudall *et al.* (2013). This character remains controversial due to the absence of reviews in several clades. In hornworts and mosses, stomata originate from a single epidermal cell, and are therefore perigenous. Angiosperms largely coded as unknown, due to a lack of information for specific taxa within the framework of Rudall *et al.* (2013). There are many reviews of angiosperms, but these are likely unreliable and taxa were scored in a general approach following the reconstruction of Rudall *et al.* (2013). Early divergent angiosperms were coded according to Rudall and Knowles (2013). The inferred ancestral condition was used for eudicots and monocots. *Oryza sativa* was scored mesoperigenous as clarified by Rudall *et al.* (2013). Fossils are scored unknown, apart from Bennettitales, which are perigenous (Rudall *et al.* 2013).

61 – 108. ***Shoot anatomy***

61. (M92) **MOD ~~Xylem~~ Internal strand of water conducting cells: Absent (0); Present (1).**
This character was modified from xylem, to include water-conducting cells without secondary wall thickenings within the sporophyte of some bryoid mosses & *Takakia* (hydroids) (Ligrone *et al.* 2000), and some extinct land plants (Edwards *et al.* 2003). These internal water-conducting cells are not present in hornwort sporophytes and are only present in the gametophytes of liverworts, hence scored absent. Water-conducting cells are present in all extant tracheophytes. Unless preservation allows, scored unknown for extinct taxa.

62. NEW Thickness of primary water-conducting cell walls: Thin (0); Thick (1).
Rule 30*. Contingent on the presence of water-conducting cells (character 61, state 1).*
All taxa without an internal strand of water-conducting cells are scored inapplicable. The water-conducting cells of *Takakia* and brylean mosses are uniformly thin-walled (Ligrone *et al.* 2000). The walls of some Polytrichales are unevenly thickened. All extant tracheophytes possess thick-walled water-conducting cell walls. Unless preservation allows, scored unknown for extinct taxa.

63. NEW Perforate primary walls of water-conducting cells: Absent (0); Present (1).
Rule 31*. Contingent on the presence of water-conducting cells (character 61, state 1).*
All taxa without an internal strand of water-conducting cells are scored inapplicable. In *Takakia*, unlike other mosses, the primary water-conducting cell walls are perforated with plasmodesmata-derived pores (Ligrone *et al.* 2000).

64. NEW Pitting size in the primary walls of perforate water-conducting cells: Small (0); Large (1).
Rule 32*. Contingent on the presence of perforate water-conducting cells (character 63, state 1).*
All taxa without a perforated internal strand of water-conducting cells are scored inapplicable. The pores in *Takakia* are small (Ligrone *et al.* 2000).

65. MOD (KC3.7, KC4.17, KC5.9) **Tracheary elements: Absent (0); Present (1).**
Rule 33*. Contingent on the presence of water-conducting cells (character 61, state 1).*
Rhyniophytes and other early land plants are only scored present if there is explicit evidence for tracheids (thick primary walls and secondary thickenings). *Aglaophyton* and *Horneophyton* are scored absent, as these taxa have unthickened hydroids (Doyle 2013); although see Ligrone *et al.* (2000) on discussion about WCCs of *Aglaophyton*). Otherwise all land plants, including fossils, are scored present, with the exception of *Ceratophyllum* ((Doyle and Endress 2010).

67. MOD (KC4.18, KC6.29, S53, HB22, HX13) **Primary tracheid ornamentation: Helical/annular/G-type (0); Scalariform/Reticulate (1); Circular bordered pits (2); Conifer type pits (3).**
Rule 34*. Contingent on the presence of tracheary elements (character 65, state 1).*
All taxa without tracheary elements in internal water-conducting strands were scored in applicable. The primary xylem includes both protoxylem and metaxylem, and scoring of character 22 in Hilton and Bateman (2006) and character 53 in Schneider *et al.* (2009) was included here. Unless preservation allows, extinct taxa are scored unknown. Schneider *et al.* (2009) scored most angiosperms as having circular bordered pits. Primary xylem in angiosperms is complex, often showing a gradual change from helical thickenings to bordered pits (Bierhorst and Zamora 1965). All angiosperms are scored as possessing bordered pits, though it was noted that taxa lacking vessels (Nymphaeales and other early groups) possess scalariform walls (Carlquist and Schneider 2007).

68. MOD (KC 4.19, HX14) **Simple pitlets in tracheid lateral wall: Absent (0); Present (1).**
Rule 35*. Contingent on the presence of tracheary elements (character 65, state 1).*
Irregular pit-like openings in the cell wall, found between annular and helical bars in zosterophylls and early lycopsids. This character is only applicable to those taxa with tracheids with secondary wall thickenings.

69. (HB27) **Torus-margo in tracheid bordered pits: Absent (0); Present (1).**
Rule 36*. Contingent on circular bordered pits (character 67, states 2, 3).*
Torus-margo refers to the thickening of the centre of the pit membrane, thus only applicable to plants with tracheids constructed from circular bordered pits. This feature is common in conifers and Gnetophyta.
Scored inapplicable in most fossil taxa, though in progymnosperms and pteridosperms where it has not been observed it is scored unknown. All angiosperms are scored absent. Outside gymnosperms, this character is only found in *Botrychium*, though it is unknown if they are homologous (Rothwell and Karrfalt 2008).

70. NEW Internal food-conducting cells: Absent (0); Present (1).
Internal food-conducting cells include phloem found in vascular plants, leptoids found in Polytrichaceae mosses and the specialised parenchyma cells in non-Polytrichaeae mosses (Ligrone *et al.* 2000). Food-conducting cells are not known in any green algae. They are absent in the sporophytes of liverworts or hornworts. Most moss sporophytes do possess internal food-conducting cells, apart from *Sphagnum* (Ligrone and Duckett 1998). Unless preservation allows identification of food-conducting cells, scored unknown for extinct taxa.

71. NEW Microtubular cytoskeleton of food conducting cells: Actin-dominant (0), Tubulin-dominant (1).
Rule 37*. Contingent on the presence of food conducting cells (character 70, state 1).*
All bryophytes with water-conducting cells are scored tubulin-dominant, while the food-conducting cells of vascular plants are actin-dominant. Extinct taxa are scored unknown.

72. (M93) **Phloem: Absent (0); Present (1).**
Rule 38*. Contingent on the presence of food conducting cells (character 70, state 1).*
Phloem is here considered a trait shared among tracheophytes (Doyle 2013), while polytrichaceae mosses possess a distinct conducting elements (leptoids), and so only tracheophytes are scored present. Extinct taxa lacking explicit description of a phloem are scored unknown. Though not always explicit, the presence of phloem in lycopsids and other tracheophytes was assumed.

73. MOD (DE16) **Sieve element plastids: S-type (0); P-I type (1); P-II type (2); no inclusion (3).**
Rule 39*. Contingent on the presence of sieve elements in phloem (character 72, state 1).*
Sieve cells are only present in phloem, therefore inapplicable to mosses. P-type plastids are restricted to seed plants. Gymnosperms were scored according to (Behnke 1974), including *Gnetum* and *Ephedra*. This character is poorly described in lower vascular plants, though *Psilotum* and *Lycopodium* possess S-type. *Selaginella*, *Isoetes* and *Equisetum* rarely or never possess starch, so have neither S nor P type, and so are scored as no inclusion (3). The remaining vascular plants and extinct taxa are scored unknown.

74. MOD (S55) **Food conducting cells (sieve elements or leptoids) with refractive spherules: Absent (0); Present (1).**
Rule 40*. Contingent on the presence of food-conducting cells (character 70, state 1).*
Refractive spherules appear to be common in euphyllophytes but absent in lycophytes. Schneider *et al.* (2009) scored all seed plants as present, yet there are no explicit references for this, and Evert (2006) describes them only as a feature of

vascular cryptograms. It is thus uncertain whether they are present among seed plants and for now will be scored unknown. Reflective spherules are present in polytrichales mosses. Fossil taxa are scored unknown.

75. (HB29) **Companion cells in phloem: Absent** (**0**)**; Present** (**1**)**.**
Rule 41*. Contingent on the presence of phloem (character 72, state 1).*
Companion cells are known only in the phloem of angiosperms (Schneider *et al.* 2009). Fossil taxa lacking clear preservation are scored unknown.

76. NEW Pericycle: Absent (**0**)**; Present** (**1**)**.**
Rule 42*. Contingent on the presence of phloem (character 72, state 1).*
Pericycle refers to the layer(s) of non-vascular cells between the between the endodermis and the phloem. Present in extant tracheophytes. Fossil taxa lacking clear preservation are scored unknown. Not observed in some early rhyniophytes e.g. *Horneophyton, Aglaophyton.*

77. (DE17) **Fibers or sclerenchyma in the pericyclic area: Present** (**0**)**; Absent** (**1**)**.**
Rule 43*. Contingent on the presence of a pericycle (character 76, state 1).*
There is no information on this for taxa beyond angiosperms, but the character was maintained as a useful character for angiosperms, with all other vascular plants scored unknown.

78. (DE18) **Pericyclic ring: Separate fiber bundles with no intervening fibers or sclerenchyma** (**0**)**; More or less continuous ring of fibers and non-hippocrepiform sclereids** (**1**)**; More or less continuous ring of fibers and hippocrepiform sclereids** (**2**)**; Homogenous ring of fibers** (**3**)**.**
Rule 44*. Contingent on the presence of fibers or sclerenchyma in pericycle (character 77, state 1).*
As with character 77, there is no information for taxa beyond angiosperms. Currently all other taxa scored as unknown.

79. MOD (S51) **Endodermal cells in roots and shoots: Primary type** (**0**)**; Secondary type** (**1**)**; Tertiary type** (**2**)**; Absent** (**3**)**.**
Rule 45*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
The endodermis, a layer of tissue surrounding the vascular bundle in tracheophytes, is absent from bryophytes. The endodermal cells of all tracheophytes, show a similar development. Initially, a Casparian strip is formed (primary), followed by a suberin lamella (secondary). Only in angiosperms does the cell wall thickness increase in an additional step (tertiary) (Schneider *et al.* 2009). Not observed in extinct protracheophytes where anatomy is preserved (i.e. Rhynie Chert plants); otherwise extinct taxa are scored unknown due to poor preservation.

80. MOD (DE2, HB18, S26, M107) **Stele type: Eustele** (**0**)**; Protostele** (**1**)**; Amphiphloic siphonostele/solenostele** (**2**)**; Ectophloic siphonostele** (**3**)**; Dictyostele** (**4**)**; Atactostele** (**5**)**.**
Rule 46*. Contingent on the presence of phloem (character 72, state 1).*
The stele is a strand of the vascular bundle (xylem/ water-conducting cells and phloem) within tracheophytes, thus those taxa lacking phloem are scored inapplicable (algae and bryophytes). The eustele (0) includes pseudosiphonosteles and reduced eusteles. Basal embryophytes scored according to descriptions in Kenrick and Crane (1997). Extinct taxa lacking clear preservation are scored unknown.

81. MOD (KC 4.14, KC5.7, KC6.23, G8, HB18, HX11) **Xylem strand shape in protostele: Terete (0); Elliptical (1); Stellate (2); Lobed (3); Clepsydroid (4); Medullated (P-type) (5); I-type (6).**
Rule 47*. Contingent on the presence of a protostele (character 80, state 1).*
This character only applies to tracheophyte taxa with a protostele.

82. COMB (KC4.15, KC5.8, KC6.24, HB16, G10, HX12) **Protoxylem position: Centrarch (0); One or more sympodia (1).**
Rule 48*. Contingent on the presence of tracheary elements (character 65, state 1).*
This character is only applicable to plants with xylem; thus algae and bryophytes scored inapplicable. Ferns scored according to Schneider *et al.* (2009). All extant seed plants scored as having one or more sympodia. Fossil taxa without preserved vascular tissue are scored unknown.

83. COMB (S24, KC4.15, KC5.8, KC6.24, G10, HX12) **Sympodia development: Exarch (0); Mesarch (1); Endarch (2).**
Rule 49*. Contingent on the presence of sympodia (character 82, state 1).*
Taxa lacking protoxylem or with centarch protoxylem are scored inapplicable. Conifers and Gnetales are generally characterised by endarch development (Taylor *et al.* 2009), and additional fossil seed plants are scored according to descriptions of Taylor *et al.* (2009). and references therein. Several taxa lacked clear information and are scored unknown. Angiosperms are scored as possessing endarch protoxylem.

84. (DE4) **Protoxylem lacunae: Absent (0); Present (1).**
Rule 50*. Contingent on the presence of tracheary elements (character 65, state 1).*
Protoxylem lacunae are known only in angiosperms and few additional taxa, such as *Equisetum, Rhacophyton* and *Ibyka* (Kenrick and Crane 1997), though it is not certain that the structures are homologous. Fossil taxa are mostly scored unknown.

85. (KC4.20, HX16) **Aligned xylem: Absent (0); Present (1).**
Rule 51*. Contingent on the presence of tracheary elements (character 65, state 1).*
The radially aligned metaxylem of basal euphyllophytes and arborescent lycophytes differs fundamentally from the bifacial cambium arrangement of seed plants (Kenrick and Crane 1997). The arborescent lycopsid *Paralycopodites* was scored present.

86. (S27) **Vascular stele cycles: Monocyclic (0); Polycyclic (1).**
Rule 52*. Contingent on the presence of tracheary elements (character 65, state 1).*
Polycyclic stems were deemed a unique feature of some ferns (Schneider *et al.* 2009).

87. (KC6.25) **Stelar suspension: Absent (0); Stelar cavity (1); Partial cavity (2).**
Rule 53*. Contingent on the presence of tracheary elements (character 65, state 1).*
This trait is a feature of extant Sellaginaceae (Kenrick and Crane 1997). Apart from exceptionally preserved examples, fossils are scored unknown.

88. COMB (KC6.28, HB2, DE6, M106, S28) **Secondary growth via vascular cambium: Absent (0); Present (1).**
Rule 54*. Contingent on the presence of water-conducting cells (character 61, state 1).*
Scoring of secondary growth in angiosperms follwing Doyle and Endress (2010). As the presence of a vascular cambium in *Glaucidium* is uncertain, it is scored unknown.

89. MOD (DE15) **Secondary phloem: Simple** (**0**)**; Stratified, fibers in tangential rows** (**1**)**;** <u>**Absent** (**2**)</u>**.**
Rule 55*. Contingent on the presence of secondary growth (character 88, state 1).*
Not applicable to plants lacking secondary growth. *Cordaixylon* and *Pentaxylon* lack fibers, but their secondary phloem is arranged in tangential bands. Extant conifers and Gnetales have tangential rows of fibers. Species for which information is lacking or unclear are scored unknown. Fossil arborescent lycopsids had unifacial cambium and thus only produced secondary xylem.

90. (KC4.21) **Xylem rays: Absent** (**0**)**; Present** (**1**)**.**
Rule 56*. Contingent on the presence of secondary growth (character 88, state 1).*
Xylem rays develop in secondary xylem. Extant taxa scored based on characters DE12 and HB28. Rays are not known in ferns and lycophytes, though *Botrychium* has rows of ray-like parenchyma (Rothwell and Karrfalt 2008).

91. COMB (DE12, HB28) **Xylem ray type: Narrow, uniseriate** (**0**)**; Wide, multiseriate** (**1**)**.**
Rule 57*. Contingent on the presence of xylem rays (character 90, state 1).*
All taxa lacking rays were scored not applicable; *Botrychium* was scored unknown.

92. (DE7) **Storied structure in tracheids, phloem and axial parenchyma: Absent** (**0**)**; Present** (**1**)**.**
Rule 58*. Contingent on the presence of secondary growth (character 88, state 1).*
Not applicable in taxa lacking secondary growth. Scored absent in extant non-angiosperms.

93. (HB23) **Secondary xylem tracheids: Circular bordered pits or perforations only** (**0**)**; At least some scalariform pitting** (**1**)**.**
Rule 59*. Contingent on the presence of secondary growth (character 88, state 1).*
Detailed information is not available for many angiosperms and so they are scored unknown. Gymnosperms tend to exclusively possess circular bordered pits, whereas scalariform pitting is more common in angiosperms. Many derived angiosperms have fibers instead of tracheids – these possess reduced pits and perforations. Additionally, taxa were scored as follows: *Ascarina* and *Hedyosmum* scalariform (Carlquist 1990), Cannellaceae scalariform (Wilson 1965), *Gomortega* circular bordered pits (Stern 1955), *Idiospermum* has bordered pits (Wilson 1979), Papaveraceae has reduced pits (Carlquist and Zona 1988), *Degeneria*, Lazardibalaceae, and Menispermaceae have fully-bordered fiber tracheids (Carlquist 1984b, 1989, 1996a), Atherospermataceae, Monimiaceae and Proteaceae have bordered pits (Patel 1973, 1992), *Amborella*, Schisandraceae, *Sarcandra* and *Trimenia* have some scalariform, though *Illicium* has fully bordered pits (Carlquist 1984a, 1999). Scalariform pitting was also present in Magnolideae (Kedrov and Timonin 2013), Saururaceae (Schneider and Carlquist 2001), *Trochodendron* and *Tetracentron* (Thompson and Bailey 1916), *Liriodendron* (Jarmolenko 1939). As is apparent, a comprehensive review of this character is required. Taxa lacking secondary growth scored inapplicable.

94. (HB24) **Tertiary spiral thickening of tracheids: Absent** (**0**)**; Present** (**1**)**.**
Rule 60*. Contingent on the presence of secondary growth (character 88, state 1).*
This is a feature of Cephalotaxus and Taxaceae (Doyle 1996), with all other taxa being scored absent.

95. NEW Axial xylem parenchyma: Absent or scarce (**0**)**; Prolific** (**1**)**.**
Rule 61*. Contingent on the presence of secondary growth (character 88, state 1).*

96. (DE13) **Paratracheal parenchyma: Absent/Scanty** (**0**)**; Well developed** (**1**)**.**
Rule 62*. Contingent on the presence of axial xylem parenchyma (character 95, state 1).*

This character is only present in parenchyma of angiosperm hardwoods (Carlquist 2001), yet is found in *Gnetum.*

97. (DE14) **Apotracheal parenchyma bands: Absent** (**0**)**; Present** (**1**)**.**
Rule 63*. Contingent on the presence of axial xylem parenchyma (character 95, state 1).*
This character is only present in parenchyma of angiosperm hardwood. Axial parenchyma in *Gnetum* and *Welwitschia* is described as scattered and diffuse (Carlquist 1996b).

98. (HB25) **Vessels: Absent** (**0**)**; Present** (**1**)**.**
Rule 64*. Contingent on the presence of tracheary elements (character 65, state 1).*
Following the work (Carlquist and Schneider 2001, 2007), it is now apparent that vessels are present in a wider variety of taxa than was previously thought. Though data is not available for all taxa, it is now know that vessels are present in the Gnetales and many groups of ferns. It is not certain how these structures relate to those of angiosperms, though it is likely that they are not homologous. Early diverging angiosperms, including the Nymphaeales, lack vessels.

99. MOD (DE9) **Vessel plate / End wall perforations: Scalariform** (**0**)**; Scalariform and simple in the same wood** (**1**)**; Simple** (**2**)**; <u>Foraminate</u> (<u>3</u>).**
Rule 65*. Contingent on the presence of vessels (character 98, state 1).*
This character was adjusted from the original to accommodate the foraminate state in *Ephedra.* Doyle and Endress (2010) also considered this character homologous to end-wall pits in vessel-less taxa. Ferns and Gnetales were scored according to Carlquist and Schneider (2001).

100. (DE11) **Vessel grouping: Predominantly solitary** (**0**)**; Mostly paired or grouped** (**1**)**.**
Rule 66*. Contingent on the presence of vessels (character 98, state 1).*
Taxa lacking vessels scored not applicable. No information was provided for ferns or *Welwitschia.*

101. (DE3) **Inverted cortical bundles: Absent** (**0**)**; Present** (**1**)**.**
Rule 67*. Contingent on the presence of water-conducting cells (character 61, state 1).*
Only present in *Idiospermum* and *Calycanthoideae*, so assumed to be a uniquely derived trait of these taxa and so scored absent in all other taxa, except fossils, which were scored as unknown.

102. MOD (DE5) **Pith: Uniform parenchymatous** (**0**)**; Septate sclerenchyma** (**1**)**; Sclerenchyma, not septate** (**2**)**; <u>Central pith canal /cavity</u> (<u>3</u>).**
Rule 68*. Contingent on the presence of a siphonostele or eustele (character 80, states 0, 2, 3, 4, 5).*
Pith is only found in eusteles and siphonosteles, i.e. taxa with protosteles were scored not applicable. This character was expanded due to the presence of sclereids in the pith of many gymnosperms (heterocellular pith), but not arranged in septa. *Cordaites*, *Emporia* and *Barthel*ia possess septate sclerenchyma and are distinct from other gymnosperms (Taylor *et al.* 2009). Several gymnosperms including *Pinus, Ginkgo, Gnetum, Thucydia* and *Ephedra* have a parenchymatous pith (Carlquist 1992), but some members of Pinaceae, Taxaceae, *Cephalotaxus* and *Podocarpus* have sclereids (Doyle and Doyle 1948; Rao and Malaviya 1964; Smith and Stockey 2001). Among *Cordaitales, Shanxioxylon* and *Cordaixylon* have septate plates, but *Mesoxylon* has a wholly parenchymatous pith (Taylor *et al.* 2009). *Lyginopteris, Pentoxylon* and *Medullosa* have sclerotic nests, yet *Pitus* and *Callistophyton* has a parenchymatous pith (Taylor *et al.* 2009). Information in ferns is scant, yet the pith is lost in *Equisetum* and has been converted to sclerenchyma in *Psilotum* (Kramer and Green 1990). Blechnaceae have some sclereids, yet the tree ferns *Dicksonia* and *Cyathea* are parenchymatous (Kramer and Green 1990).

103. NEW Fibers: Absent (0); Present (1).
Rule 69*. Contingent on the presence of water-conducting cells (character 61, state 1).*
Fibers and fiber tracheids are found only in angiosperms and Gnetales. Non-lignified fibers occur in *Gnetum*.

104. (DE10) **Fiber pitting (lateral pitting of tracheids in vesselless taxa): Distinctly bordered (0); Minutely bordered (1).**
Rule 70*. Contingent on the presence of fibers (character 103, state 1).*
Fibers and fiber tracheids are found only in angiosperms and Gnetales. All other taxa were scored inapplicable. The pitting in Gnetales is distinctly bordered (Carlquist 1996b).

105. (DE19) **Laticifers in stem: Absent (0); Present (1).**
Rule 71*. Contingent on the presence of a persistent sub-sporangial stalk formed by cell division (character 46, state 1).*
A feature of angiosperms, though also present in the pith of *Gnetum* (Kramer and Green 1990). All other extant taxa scored absent; extinct taxa scored unknown.

106. (DE20) **Raphide idioblasts: Absent (0); Present (1).**
Rule 72*. Contingent on the presence of a persistent sub-sporangial stalk formed by cell division (character 46, state 1).*
A feature of angiosperms (Bruni *et al.* 1982). All other extant taxa scored absent. Fossil taxa scored unknown.

107. (M103) **True lignin: Absent (0); Present (1).**
Scored according to (Weng and Chapple 2010). All higher land plants, including fossils, were scored present. This is mainly due to the ubiquity of lignin in structures like tracheids and sclerenchyma. Only lower fossil taxa where no explicit reference to these structures is made were scored unknown.

109 – 116. *Branching*

109. (KC3.4, HX3) **Sporophyte branching: Monosporangiate (0); Polysporangiate (1).**
Rule 73*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
All tracheophytes, including most fossils, were scored as polysporangiates, as a result of sporophyte branching.

110. MOD (HB4, M109, S33) **Primary branching system: Solely dichotomous via apical meristem (0); Axillary, via one or more axillary buds (1).**
Rule 74*. Contingent on the presence of sporophyte branching (polysporangiate) (character 109, state 1).*
Axillary branching is only known to occur in seed plants (Schneider et al., 2009). Monosporangiate taxa scored not applicable.

111. MOD (KC4.4, HX4) **Type of dichotomous branching system: More or less isotomous (0); Pseudomonopodial or anisotomous (1).**
Rule 75*. Contingent on a dichotomous branching system (character 110, state 0).*
Modified from the original character to include anisotomous with pseudomonopodial to accommodate the lateral branching pattern of some tracheophytes. *Lygodium, Osmunda* and *Gleichenia* show an isotomous branching pattern (Schneider 2013). Taxa that branch via axillary buds (i.e. monopodial) and monosporangiate taxa were scored not applicable.

112. MOD (KC4.5, 5.2, 6.3, HX5) **Branching pattern in dichotomous branching system: No pattern (0); Planated (1); Non-planar (2).**

Rule 76*. Contingent on a dichotomous branching system (character 110, state 0).*

Taxa with axillary branching and monosporangiate taxa were scored not applicable. Ferns which do not branch (Schneider 2013) were scored inapplicable.

113. MOD (KC4.6) **Subordinate branching in dichotomous branching system: Absent (0); Present (1).**

Rule 77*. Contingent on a dichotomous branching system (character 110, state 0).*

Taxa with axillary branching and monosporangiate taxa were scored not applicable. Kenrick and Crane (1997) defined this as a feature of some zosterophylls and extinct *Selaginella*. All other taxa scored as absent.

114. (KC4.7) **Rhynia-type adventitious branches: Absent (0); Present (1).**

Rule 78*. Contingent on a dichotomous branching system (character 110, state 0).*

Only applies to *Rhynia* and *Huvenia* (Kenrick and Crane 1997). Taxa with axillary branching or monosporangiate taxa were scored not applicable.

115. COMB (KC4.8, HX6, S3) **Circinate vernation and recurvation: Absent (0); Circinate vernation (1); Recurvation (2).**

Rule 79*. Contingent on the presence of sporophyte branching / multiple sporangia (character 109, state 1).*

Refer to the coiling of the tip of an axis or leaf bud. Scoring of Schneider *et al.* (2009) used. Only Droseraceae among angiosperms show circinate vernation, which are not included in the present study. Several fossil taxa scored unknown. Monosporangiate taxa were scored not applicable.

116. (HB5, S35) **Vegetative short shoots: Absent (0); Present (1).**

Rule 80*. Contingent on the presence of sporophyte branching / multiple sporangia (character 109, state 1).*

Restricted to some gymnosperms (Schneider *et al.* 2009), the scoring of the original character was changed to present in *Gnetum* (Schneider *et al.* 2009).

117 – 165. ***Appendages***

118. MOD (KC4.9, KC5.5) **Multicellular appendages: Absent (0); Present (1).**

Rule 81*. Contingent on the presence of a persistent sub-sporangial stalk formed by cell division (character 46, state 1).*

Restricted to enations of early land plants and lycophylls (= microphylls), while 'dichotomous pinnule-like appendages', fronds and euphylls (megaphylls) are essentially considered to be the ultimate branching units of lateral branching systems.

119. MOD (KC6.7) **Vascularisation of multicellular appendage: Absent (0); Full (1); Partial (2).**

Rule 82*. Contingent on the presence of multicellular appendages (character 118, state 1).*

Enations of early land plants are not vascularised, while lycophylls possess single, unbranched veins. Partial vascularisation of the multicellular appendages occur in *Asteroxylon*. All taxa lacking multicellular appendages as defined in character 118 were scored not applicable.

120. NEW Lycophylls: Absent (0); Present (1).

Rule 83*. Contingent on the presence of vascularisation in multicellular appendages (character 119, states 1, 2).*

Refers only the the specialised microphyllous leaves of Lycopsida. Single vein or reduced vascular tissue. Lack apical meristem, leaf primordial from few initials, absence of blade/petiole differentiation. All taxa lacking multicellular appendages as defined in character 118 were scored not applicable.

121. MOD (KC6.9) **Lycophyll leaf shape: Simple** (**0**)**; Forked** (**1**)**.**
Rule 84*. Contingent on the presence of lycophylls (character 120, state 1).*
All taxa not possessing lycophylls were scored not applicable.

122. (KC6.11) **Ligule: Absent** (**0**)**; Present** (**1**)**.**
Rule 85*. Contingent on the presence of lycophylls (character 120, state 1).*
All taxa not possessing lycophylls were scored not applicable.

123. MOD (KC6.12) **Mucilage canals in lycophylls: Absent** (**0**)**; Present** (**1**)**.**
Rule 86*. Contingent on the presence of lycophylls (character 120, state 1).*
All taxa not possessing lycophylls were scored not applicable.

124. (KC6.5) **Bulbils: Absent** (**0**)**; Present** (**1**)**.**
Rule 87*. Contingent on the presence of lycophylls (character 120, state 1).*
Bulbils (gemmae) are a characteristic feature of some homosporous lycophytes (Kenrick and Crane 1997). All other taxa with multicellular appendages and clear preservation were scored unknown. All taxa not possessing lycophylls were scored not applicable.

125. MOD (KC6.10) **Lycophyll anisophylly: Absent** (**0**)**; Present** (**1**)**.**
Rule 88*. Contingent on the presence of lycophylls (character 120, state 1).*
All taxa not possessing lycophylls were scored not applicable.

126. MOD (KC5.6, KC6.8) **Phyllotaxy of multicellular appendages: Absent** (**0**)**; Spiral** (**1**)**; Two-rowed** (**2**)**; Four-rowed** (**3**)**; Opposite / Whorled** (**4**)**; Distichous** (**5**)**.**
Rule 89*. Contingent on the presence of multicellular appendages (character 118, state 1).*
All taxa not possessing multicellular appendages, as defined in character 118, were scored not applicable.

127. MOD (KC4.10, HX8) **Dichotomous pinnule-like multicellular appendages: Absent** (**0**)**; Present** (**1**)**.**
Rule 90*. Contingent on the presence of a persistent sub-sporangial stalk formed by cell division (character 46, state 1).*
"Dichotomous pinnule-like ultimate appendages" of Kenrick and Crane (1997). Enations of early land plants and lycophylls (= microphylls) are unbranched; dichotomous pinnule-like appendages, including 'ultimate branching systems', fronds and euphylls (=megaphylls), are branched. Scored separately to unbranched multicellular appendage as some early land plants possess both enations and ultimate branching systems.

128. MOD (R18, X7) **Planation of dichotomous pinnule-like multicellular appendages: Non-planar** (**0**)**; Planar** (**1**)**.**
Rule 91*. Contingent on the presence of dichotomous pinnule-like multicellular appendages (character 127, state 1).*
Euphylls are planar dichotomous multicellular appendages, and some early lateral branching systems are planar.

129. NEW Webbing of of dichotomous pinnule-like multicellular appendages: Absent (**0**)**; Present** (**1**)**.**
Rule 92*. Contingent on the presence of dichotomous pinnule-like multicellular appendages (character 127, state 1).*
Present in euphylls; absent in UBU of LBS of the most basal euphyllophytes.

130. NEW Euphylls: Absent (0); Present (1).
Rule 93*. Contingent on the presence of dichotomous pinnule-like multicellular appendages (character 127, state 1).*
Euphylls are planar appendages with webbing, multiple branched vascular strands, leaf gaps, and development via an apical meristem.

131. COMB (KC5.6, KC6.8 , HB6, DE21, DE22, HX9) **Phyllotaxis of dichotomous pinnule-like appendages: Absent (0); Spiral (1); 2-rowed (2); 4-rowed (3); Opposite/whorled (4); Distichous (5).**
Rule 94*. Contingent on the presence of dichotomous pinnule-like multicellular appendages (character 127, state 1).*
All taxa lacking multicellular dichotomous pinnule-like appendages scored not applicable. All ferns were scored as spiral, with the exception of *Equisetum* and *Salvinia* which were whorled. Fossil taxa lacking clear preservation scored unknown.

132. (S132) **Number of photosynthetic leaves per shoot at any given time: Two or more (0); One (1).**
Rule 95*. Contingent on the presence of euphylls (character 130, state 1).*
This character was determined as an autapomorphy for Ophioglossales (Schneider *et al.* 2009). Fossil taxa lacking leaves scored unknown. Taxa lacking euphylls scored not applicable.

133. MOD (S22) **Euphyll intercalary meristem: Absent (0); Present (1).**
Rule 96*. Contingent on the presence of euphylls (character 130, state 1).*
Wording changed to distinguish it from the intercalary growth of hornwort sporophytes and the intercalary regions of *Equisetum* stems. The presence of intercalary growth in seed plants is a fundamental difference between them and other euphyllophytes (Schneider *et al.* 2009). It is present in the elongated leaves of conifers and *Welwitschia* (Beck 2010). No information was recovered for Zamiaceae nor Ephedraceae. Fossils scored unknown.

134. COMB (HB7, S5, DE30, DE34) **Leaf form: Pinnately compound (0); Simple, pinnately veined, or dissected into parallel veined segments (1); Linear to dichotomous with two or more veins (2); Palmately veined (actinodromous or acrodomomous) (3); Linear with one vein (rarely two; may fork apically, as in Emporia) (4).**
Rule 97*. Contingent on the presence of euphylls (character 130, state 1).*
All taxa lacking euphylls scored not applicable. Fossils lacking well-preserved euphylls scored unknown.

135. (DE34) **Leaf dissection: Simple (0); Compound or lobed (1).**
Rule 98*. Contingent on the presence of euphylls (character 130, state 1).*
This character provides further differentiation among palmately veined leaves, although there is some overlap with Character 138. All taxa lacking euphylls scored not applicable. Fossils lacking well preserved euphylls scored unknown.

136. (HB8) **Rachis: Bifurcate (0); Simple (1).**
Rule 99*. Contingent on the presence of pinnately compound leaves (character 135, state 1).*
This character is only applicable to taxa with pinnately compound leaves. Bifurcate fronds are characteristic of lyginopterids and some other seed ferns (Stewart and Rothwell 2010). As there is no reference to this structure in ferns, they were scored absent.

137. (S7) **Leaf architecture: Monopodial, primary rachis longer than secondary (0); Sympodial, secondary rachis longer than primary rachis (1); Sympodial, secondary rachis longer than primary rachis, dormant bud terminating the primary rachis (2).**

Rule 100*. Contingent on the presence of euphylls (character 130, state 1).*

Taxa without simple meristems or euphylls scored not applicable (Schneider *et al.* 2009). Sympodial architecture is a derived trait of the Gleicheniales. Fossils lacking clear euphylls scored unknown.

138. COMB (HB9, DE31, S10) **Leaf venation: Open (0); Reticulate (1).**

Rule 101*. Contingent on the presence of euphylls (character 130, state 1).*

All taxa lacking euphylls scored not applicable. Fossils lacking clear euphylls scored unknown.

139. COMB (HB10, S9) **Vein orders: One (0); Two or more (1).**

Rule 102*. Contingent on the presence of euphylls (character 130, state 1).*

All angiosperms were assumed to have two or more orders of venation (Melville 1969). Taxa lacking euphylls scored not applicable. Fossils lacking clear leaves scored unknown.

140. (S8) **Primary blade form: Solitary/unbranched (0); Dichotomous/isotomous (1); Non-dichotomous/anisotomous (2).**

Rule 103*. Contingent on the presence of euphylls (character 130, state 1).*

Taxa lacking euphylls scored not applicable. Most conifers and *Ephedra* possess simple unbranched venation (Stevenson 2012), though some Podocarpaceae and Araucariaceae possess more complex patterns (Gifford and Foster 1989). In other cases, other characters such as pinnate venation or higher order branching were used to infer an anisotomous branching pattern. Fossils lacking clear leaves scored unknown.

141. COMB (HB13, S8) **Leaf traces: Mesarch (0); Endarch (1).**

Rule 104*. Contingent on the presence of euphylls (character 130, state 1).*

All taxa lacking euphylls scored not applicable. Endarch leaf traces represent a synapomorphy of higher seed plants, with a subsequent reversal in Cordaites (Doyle 1996), so all angiosperms were scored endarch. Fossil taxa lacking clear leaf traces were scored unknown.

142. (HB14) **Girdling leaf traces: Absent (0); Present (1).**

Rule 105*. Contingent on the presence of euphylls (character 130, state 1).*

HB define this is a feature unique to Cycads. Score absent for all other extant taxa and fossil taxa with preserved leaf traces.

143. COMB (HB20, DE21) **Node anatomy: One trace to each leaf (0); More than three traces (1); Two traces from adjacent bundles (2); Three traces (3); Leaf traces from two or more protoxylem strands or bundles over a substantial length of stem (4).**

Rule 106*. Contingent on the presence of euphylls (character 130, state 1).*

All taxa lacking euphylls scored not applicable. Ferns were scored by using character S18 “Petiole stele number”, assuming that the vasculature at the base of the petiole is equivalent to that of the node. Fossils lacking clear nodal anatomy scored unknown.

144. COMB (HB15, DE39) **Astrosclereids in leaf: Absent (0); Present (1).**

Rule 107*. Contingent on the presence of euphylls (character 130, state 1).*

Taxa lacking euphylls scored not applicable. A synapomorphy of Gnetales and some angiosperms (Doyle 1996), all other taxa scored absent. Fossils lacking clear leaves scored unknown.

145. (S4) **Fertile/sterile leaf differentiation: (nearly) Monomorphic (0); Hemidimorphic (1); Dimorphic (2).**
Rule 108*. Contingent on the presence of euphylls (character 130, state 1).*
Taxa lacking euphylls scored not applicable. All higher seed plants scored dimorphic. Taxa lacking clear leaves scored unknown.

146. (S11) **Blade scales: Absent (0); Present (1).**
Rule 109*. Contingent on the presence of euphylls (character 130, state 1).*
This character is acknowledged as not being independent of the shoot indumentum character (52-53), since taxa with blade scales also possess shoot scales. Scales are defined as not being homologous to the scales of angiosperms, which are reduced leaves (Schneider *et al.* 2009) and so were only scored present in ferns.

147. (S12) **Pulvini: Absent (0); Present (1).**
Rule 110*. Contingent on the presence of euphylls (character 130, state 1).*
This character refers only to the structures present in the Marratiaceae, rather than those found in Marsilaceae. All other taxa scored absent.

148. (S13) **Pneumathodes: Absent (0); Present and scattered all around petiole and/or rachis (1); Present and borne in discrete lines or patches on petiole and/or rachis (2).**
Rule 111*. Contingent on the presence of euphylls (character 130, state 1).*
This character refers to structures on the rachis, rather than roots, and so was scored present only in certain ferns.

149. (S17) **Epipetiolar branches: Absent (0); Present (1).**
Rule 112*. Contingent on the presence of euphylls (character 130, state 1).*
Taxa lacking euphylls scored unknown. A feature only of *Pteridium*, all other taxa with sufficient preservation scored absent.

150. (S20) **Leaf vascular bundle: Amphicribal (0); Collateral (1).**
Rule 113*. Contingent on the presence of euphylls (character 130, state 1).*
Extant taxa scored on the basis that collateral is the condition in seed plants (Beck 2010). Fossil taxa without clear anatomical detail were scored unknown. All taxa lacking euphylls were scored not applicable. The amphivasal condition of monocot vascular bundles is confined to the stem and not found in the roots, so all monocots were scored collateral (Coulter and Chamberlain 1903).

151. (DE24) **First appendages on vegetative branch: Paired prophylls (0); A single prophyll (1).**
Rule 114*. Contingent on the presence of euphylls (character 130, state 1).*
Prophylls are a synapomorphy of seed plants and so all non-seed plants were scored as not applicable (Barthelemy and Caraglio 2007). Paired prophylls are found in all conifers (Tomlinson and Zacharias 2001) and *Welwitschia* (Church 1914), but information for other non-angiosperms is scant. Given their position as a synapomorphy among seed plants, all extant seed plants were scored as having paired prophylls. Fossil seed plants for which this character has not been described were scored missing.

152. (DE25) **Leaf base: Non-sheathing** (**0**)**; Sheathing** (**1**)**.**
Rule 115*. Contingent on the presence of euphylls (character 130, state 1).*
Sheathing was defined as covering more than half the circumference of the stem. Fern fronds were scored non-sheathing, though it is noted that the leaves of *Equisetum* form a kind of basal sheath called an ochreole (Hauke (Hauke 1987) and *Ophioglossum* has a sheath (Simpson 2011). The decurrent attachment of many conifers was scored as non-sheathing. Similarly, non-sheathing leaves were found in *Thucydia*, *Pentoxylon*, *Heterangium*, *Lyginopteris*, *Archaeopteris*, in reconstructions of Bennettitales, Cordaitales, Glossopterids (Taylor *et al.* 2009), *Autunia* (Kerp 1988), *Caytonia* (Crane 1985), *Laceya* (Klavins and Matten 1996), *Elkinsia* (Serbet and Rothwell 1992) and *Cecropsis* (Stubblefield and Rothwell 1989). In contrast, the leaves of Cheirolepidiaceae (Watson 1988) and Gnetales (Loconte and Stevenson 1990) were described as decurrent and sheathing. Taxa lacking euphylls scored unknown.

153. (DE26) **Stipules: Absent** (**0**)**; Axillary/adaxial** (**1**)**; Interpetiolar** (**2**)**; Paired cap** (**3**)**.**
Rule 116*. Contingent on the presence of euphylls (character 130, state 1).*
Hickey and Wolfe (1975) describe stipules as diagnostic of angiosperms and so all other extant and fossil taxa with clear preservation were scored absent.

154. (DE27) **Axillary squamules: Absent** (**0**)**; Present** (**1**)**.**
Rule 117*. Contingent on the presence of euphylls (character 130, state 1).*
No reference to squamules are made for any taxa other than angiosperms and so all other euphyllophytes were scored absent.

155. (DE28) **Leaf blade: Bifacial** (**0**)**; Unifacial/tetragonal** (**1**)**.**
Rule 118*. Contingent on the presence of euphylls (character 130, state 1).*
Modified from the original character to incorporate the tetragonal leaf blades of some conifers (Stevenson 2012). All ferns possess bifacial leaves with the exception of *Pinnularia* (Vasco *et al.* 2013). *Cycas, Ginkgo, Taxus*, *Cephalotaxus, Welwitschia, Podocarpus* and *Gnetum* are bifacial, though many pines, Araucariaceae and *Ephedra* are tetragonal (Stevenson 2012), Döhren, 2013). The Bennettitales and Glossopterids show bifacial leaves (Pigg and McLoughlin 1997; Ray *et al.* 2014). Fossil taxa for which the leaves were not described were scored unknown.

156. (DE29) **Leaf shape: Obovate to oblong to elliptical** (**0**)**; Ovate** (**1**)**; Linear to strap-shaped** (**2**)**; Acicular/subulate** (**3**)**; Reduced/scales** (**4**)**; Flabelliform** (**5**)**; Pinnate to pinnatifid** (**6**)**.**
Rule 119*.Contingent on the presence of euphylls (character 130, state 1).*
This character was expanded from the original to include conifer and gymnosperm morphology. Ferns scored according to (Kramer and Green 1990), progymnosperms, pteridosperms and gymnosperms scored according to Taylor *et al.* (2009), Gifford and Foster (1989) and (Simpson 2011). All taxa lacking euphylls scored not applicable. Fossils lacking clearly preserved leaves scored missing.

157. (DE32) **Base of blade: Not peltate** (**0**)**; Peltate in some or all leaves.**
Rule 120*. Contingent on the presence of euphylls (character 130, state 1).*
Peltate leaf bases were not present in any non-angiosperms. All taxa lacking euphylls scored not applicable. Fossils lacking clearly preserved leaves scored missing.

158. (S6) **Dromy at base of blade: Catadromous** (**0**)**; Anadromous** (**1**)**; Isodromous** (**2**)**.**
Rule 121*. Contingent on the presence of euphylls (character 130, state 1).*

This character could only apply to taxa with pinnate leaves and is poorly documented beyond ferns. Consequently all taxa lacking pinnate leaves or euphylls were scored not applicable. Non-ferns with pinnate leaves were scored unknown.

159. (DE33) **Apex of blade: Simple** (**0**)**; Bilobed** (**1**)**.**
Rule 122*. Contingent on the presence of euphylls (character 130, state 1).*
The only taxa with bilobed leaf apices were *Liriodendron* and *Ginkgo.* All other taxa with clearly preserved leaves scored absent. All taxa lacking euphylls scored not applicable. Fossils lacking clearly preserved leaves scored missing.

160. MOD (DE35) **Marginal teeth: Absent** (**0**)**; Present** (**1**)**.**
Rule 123*. Contingent on the presence of euphylls (character 130, state 1).*
The character states describing marginal teeth in DE35 could not be determined for lower taxa possessing marginal teeth, so a new character was designed. All taxa scored for DE35 were scored present. In addition, many ferns possess teeth (Kramer and Green 1990) as well as many fossil taxa (Taylor *et al.* 2009). All taxa lacking euphylls scored not applicable. Fossils lacking clearly preserved leaves scored missing.

161. (DE35) **Marginal teeth type: Chloranthoid** (**0**)**; Monimioid** (**1**)**; Platanoid** (**2**)**.**
Rule 124*. Contingent on the presence of marginal teeth (character 160, state 1).*
Information for non-angiosperms was not available. All taxa lacking marginal teeth were scored not applicable. Non-angiosperm taxa with marginal teeth were scored unknown.

162. (DE37) **Midrib vasculature: Simple arc** (**0**)**; Arc with adaxial plate** (**1**)**; Ring** (**2**)**.**
Rule 125*. Contingent on the presence of euphylls (character 130, state 1).*
Information for this character is lacking in non-angiosperms, but was included as a valid character for angiosperms. All non-angiosperm taxa were scored unknown.

163. (DE38) **Palisade parenchyma: Absent, mesophyll homogeneous** (**0**)**; Present, mesophyll dorsiventral** (**1**)**.**
Rule 126*. Contingent on the presence of euphylls (character 130, state 1).*
Though the absence of palisade parenchyma in ferns is considered rare (Vasco *et al.* 2013), it is noted as absent in *Hymenophyllum, Psilotum* and *Tmesipteris* and other authors consider it largely absent (Lommasson and Young 1971). Consequently, it was scored as unknown in ferns were no explicit description is given. *Pteridium* (Halarewicz 2008), *Blechum* (Gloss, 1887), *Cyathea, Dicksonia, Plagiogyria* and *Marsilea* (Warmbrodt and Evert 1978, 1979b, 1979a) were scored as present. Additionally, *Lyginopteris, Heterangium*, Glossopterids, Corystosperms, Cordaitales, Caytoniales, Medullosales and Bennettitales possess palisade parenchyma, though it is likely absent in *Callistophyton, Emporia* and *Thucydia* (Harris 1940; Hernandez-Castillo *et al.* 2001; Taylor *et al.* 2009; Ray *et al.* 2014; Raymond *et al.* 2014). Among extant gymnosperms, palisade parenchyma is found in Araucariaceae, Taxaceae, Ginkgo, Cycads, *Welwitschia* and *Ephedra* (Bhatnagar and Moitra 1996; Beck 2010; Ghimire *et al.* 2014) Wang et al., 2014. Podocarpaceae, Pinaceae and Cupressaceae are reported as possessing and not possessing palisade (Larsen 1927) (Bhatnagar and Moitra 1996) Dörken 2013. Taxa lacking euphylls were scored not applicable. Fossil taxa lacking clear preservation were scored missing.

164. (DE40) **Oil cells in mesophyll: Absent** (**0**)**; Present** (**1**)**.**
Rule 127*. Contingent on the presence of euphylls (character 130, state 1).*

A recent review of oil bodies by (Lersten *et al.* 2006) looks at the phylogenetic distribution of oil bodies in mesophyll in plants, but did not agree with the original scoring. The original scoring was maintained, though non-angiosperms were scored unknown.

165. (DE41) **Mucilage cells in mesophyll: Absent** (**0**)**; Present** (**1**)**.**
Rule 128*. Contingent on the presence of euphylls (character 130, state 1).*
Despite little information for non-angiosperm taxa, this character was maintained as a valid character for angiosperms, with other taxa possessing true leaves scored unknown.

166 – 173. ***Sporophylls***

166. COMB (HB45, KC6.15) **Sporophyll: Absent** (**0**)**; Sporophyll with a single sporangium** (**1**)**; Sporophyll with two meio/microsporangia** (**2**)**; Sporophyll with more than two meio/microsporangia** (**3**)**.**
Rule 129*. Contingent on the presence of multicellular appendages (character 118, state 1).*
Ferns and angiosperms were scored as possessing more than two sporangia per sporophyll, with the exception of *Psilotum* (Schneider 2013). Fossils lacking clear preservation of the sporangia were scored unknown.

167. MOD (KC6.16) **Shape of sporophyll bearing single sporangium: More or less unmodified** (**0**)**; Subpeltate** (**1**)**.**
Rule 130*. Contingent on the presence of a sporophyll with a single sporangium (character 166, state 1).*
This character applies only to the lycopsids (Kenrick and Crane 1997).

168. MOD (KC6.19) **Structure of sporophyll bearing single sporangium: Simple** (**0**)**; Ephemeral** (**1**)**; Complex** (**2**)**.**
Rule 131*. Contingent on the presence of a sporophyll with a single sporangium (character 166, state 1).*
This character applies only to the lycopsids (Kenrick and Crane 1997).

169. (HB42) **Microsporophylls: Pinnate or paddle-like** (**0**)**; Simple, one-veined, scale-like** (**1**)**; Simple, one** (**rarely three**) **veined, with two pairs of longitudinal microsporangia** (**2**)**.**
Rule 132*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and sporophylls (character 166, states 1, 2, 3).*
All homosporous taxa and taxa lacking sporangia scored not applicable. Following the coding of Hilton and Bateman (2006), all angiosperms were scored (2).

170. (HB48) **Microsporophyll fusing: Free** (**0**)**; Basally fused** (**1**)**.**
Rule 133*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and sporophylls (character 166, states 1, 2, 3).*
All homosporous taxa and taxa lacking sporangia scored not applicable. Following the coding of Hilton and Bateman (2006) all angiosperms were scored free.

171. (HB51) **Fertile appendages: Not aggregated or in simple strobili** (**0**)**; Simple male, compound female strobili** (**1**)**; Compound male and female strobili** (**2**)**.**
Rule 134*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1).*
All homosporous taxa and taxa lacking sporangia scored not applicable. All non-seed plants were scored as having not aggregated or simple strobili (Hilton and Bateman 2006). Angiosperms were coded according to character 43 of (Doyle and Endress 2010).

172. (HB54) **Bract/shoot complexes: Absent** (**0**)**; Helical** (**1**)**; Vertical rows** (**2**)**.**
Rule 135*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1).*
Helical and vertical rows are features of conifers, Cordaitales and Gnetales (Hilton and Bateman 2006), all remaining taxa were scored absent.

173. (HB55) **Bract and axillary female shoot: Free** (**0**)**; Fused** (**1**)**.**
Rule 136*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1).*
All homosporous taxa and taxa lacking sporangia scored not applicable. Doyle (1996) argued that scoring of taxa beyond conifers raises issues of homology, so all other taxa were scored unknown.

174 – 186. ***Sporangia (general)***

174. (KC3.8) **Sporangium: Absent** (**0**)**; Present** (**1**)**.**
Sporangia are scored present in all embryophytes (Kenrick and Crane 1997).

175. (KC4.11) **Conical sporangium emergences: Absent** (**0**)**; Present** (**1**)**.**
Rule 137*. Contingent on the presence of sporangia (character 174, state 1).*
All taxa lacking sporangia were scored not applicable. These sporangia are found only in some rhyniophytes, such as *Caia* and *Horneophyton* (Kenrick and Crane 1997).

176. (KC3.10, KC4.32) **Columella: Absent** (**0**)**; Present** (**1**)**.**
Rule 138*. Contingent on the presence of sporangia (character 174, state 1).*
All taxa lacking sporangia were scored not applicable. Defined as a columnar mass of sterile tissue that develops within the spore masses of hornworts, mosses and some early polysporangiates (Kenrick and Crane 1997), all derived taxa were scored absent.

177. NEW Sporogenous tissue continuous over columella: Absent (**0**)**; Present** (**1**)**.**
Rule 139*. Contingent on the presence of a columella (character 176, state 1).*
In some bryophytes the columella occurs throughout the entire sporangial length; in others, such as *Sphagnum, Andreaea,* the columella is overtopped with sporogenous tissue.

178. (KC4.26) **Sporangium abscission: Absent** (**0**)**; Present** (**1**)**.**
Rule 140*. Contingent on the presence of sporangia (character 174, state 1).*
With the exception of taxa scored present by KC and *Zosterophyllum deciduum* (Gerrienne 1988), all other taxa were scored absent (Kenrick & Crane, 1997).

179. (S72) **Sporangia wall thickness/development: Two or more layers** (**0**)**; One cell layer** (**1**)**.**
Rule 141*. Contingent on the presence of sporangia (character 174, state 1).*

180. (M99) **Peristome: Absent** (**0**)**; Present** (**1**)**.**
Rule 142*. Contingent on the presence of sporangia (character 174, state 1).*
This structure is found only in certain species of moss, and so all other taxa were scored absent.

181. MOD (N35) **Peristome type: nematodontous** (**0**)**; haplolepidous** (**1**)**; diplolepidous** (**2**)**.**
Rule 143*. Contingent on the presence of a peristome (character 180, state 1).*
After Newton *et al.* (2000).

182. (KC4.24) **Specialised fertile zone: Absent** (**0**)**; Densely branched sporangium clusters** (**1**)**.**
Rule 144*. Contingent on the presence of sporangia (character 174, state 1) and branching (character 109, state 1).*
Taxa lacking sporangia or monosporangiates scored not applicable. Present in many progymnosperms, trimerophytes and early fern-like taxa (Kenrick and Crane 1997), this character is absent in gymnosperms and angiosperms. *Archaeopteris* is scored present, though *Cecropsis* was scored absent based on the description of Stubblefield and Rothwell (1989).

183. (S69) **Sporangial fusion resulting in a synangium: Absent** (**0**)**; Wholly or partially fused** (**1**)**.**
Rule 145*. Contingent on the presence of sporangia (character 174, state 1).*
Fused microsporangia and megasporangia are included here, as in (Doyle 1996), and so angiosperms were scored according to the presence or absence of fused stamen as in DE67 and DE107, which would represent a partial fusing.

185. (M91) **Pseudoelaters: Absent** (**0**)**; Present** (**1**)**.**
Rule 146*. Contingent on the presence of sporangia (character 174, state 1).*
Following the coding of Mishler *et al.* (1994), these were treated as a unique feature of hornworts, and so all other taxa with a preserved gametophyte were scored absent.

186. (M86) **Elaters: Absent** (**0**)**; Present** (**1**)**.**
Rule 147*. Contingent on the presence of sporangia (character 174, state 1).*
Following the coding of Mishler *et al.* (1994), this was treated as unique feature of liverworts and so all other taxa were scored absent.

187 – 211. ***Homosporous sporangia***

187. NEW Homosporous sporangium dehiscence: Indehiscent (**0**)**; Dehiscent** (**1**)**.**
Rule 148*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and the presence of sporangia (character 174, state 1).*

188. MOD (KC3.9) **<u>Homosporous</u> sporangium dehiscence <u>type</u>: Linear** (**0**)**; Radial** (**1**)**; Operculate** (**2**)**.**
Rule 149*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and dehiscent sporangia (character 187, state 1).*
All seed plants and heterosporous taxa were scored not applicable. Mosses belonging to Bryopsida were scored as operculate (Glime 2007). Hornworts and liverworts were scored as linear, as both show dehiscence according to 2, 4 or irregular valves (Glime 2007). Ferns were all scored as linear, based on the scoring of (Schneider *et al.* 2009), who describe the orientation of dehiscence as longitudinal or transverse.

189. MOD (S79) **Orientation of homosporous sporangium dehiscence: Transverse** (**0**)**; Longitudinal** (**1**)**.**
Rule 150*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and dehiscent sporangia (character 187, state 1).*
Only taxa scored for the above character were scored. Lower land plants were scored according to descriptions of Kenrick and Crane (1997).

190. MOD (KC4.30) **Thickened homosporous sporangium valve rim: Absent (0); Present (1).**
Rule 151*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and dehiscent sporangia (character 187, state 1).*
The structure of dehiscence of leptosporangiate ferns relies on thin walled cells around the stomium (King 1944; Noblin *et al.* 2012), and so leptosporangiate ferns were scored absent. Eusporangiate ferns were scored unknown. Following Kenrick and Crane (1997), all bryophytes were scored absent.

191. (KC5.15) **Relative valve size of homosporous sporangium: Isovalvate (0); Anisovalvate (1).**
Rule 152*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and non-operculate dehiscent sporangia (character 188, states 1, 0).*
All literature for this character describes lycopsids.

192. MOD (KC4.27) **Homosporous sporangium shape: More or less fusiform (0); More or less reniform (1); More or less spatulate (2); More or less spherical (3); Axis-like (4); Elliptical (5).**
Rule 153*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and the presence of sporangia (character 174, state 1).*
Heterosporous taxa and taxa lacking sporangia were scored not applicable. Ferns were scored according to (Schneider *et al.* 2009).

193. MOD (KC4.28) **Homosporous sporangium symmetry: Radial (0); Bilateral (1).**
Rule 154*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and the presence of sporangia (character 174, state 1).*
All heterosporous taxa scored not applicable. Following Reitner et al. (2013), all euphyllophytes were scored radially symmetrical. Following (Kenrick and Crane 1997), all bryophytes were scored radial.

194. MOD (KC5.18) **Homosporous sporangial attachment: Terminal (0); Lateral, on short stalks (1); Lateral, sessile (2); Attached to special pads of tissue (3); Attached dorsal/marginal leaf surface (4); Attached to apical leaf surface (5).**
Rule 155*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and the presence of sporangia (character 174, state 1).*
Heterosporous taxa and taxa lacking sporangia were scored not applicable. Marattiales and leptosporangiate ferns have sporangia on the dorsal surface, while Ophioglossales and Psilotales have them on the apical side (Schneider 2013).

196. MOD (KC6.22) **Homosporous sporangium distribution: Extended (0); Compact (1).**
Rule 156*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and laterally attached (character 194, states 1-5).*
This character refers to whether sporangia feature over a broad or localised region of the branch. Given that fern sporangia are localised to sori, they were scored compact. All taxa with terminal sporangia were scored not applicable.

197. MOD (KC4.22) **Sporangiotaxis of homosporous sporangia: None (0); Two rows (1); Four rows (2); Helical (3).**
Rule 157*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and laterally attached (character 194, states 1-5).*
All heterosporous taxa and taxa lacking lateral sporangia scored not applicable.

198. MOD (KC5.11) **Homogenous sporangia rows: Ventral (0); Dorsiventral (1); Lateral (2).**
Rule 158*. Contingent on the presence of homosporous sporogenesis (character 4, state 0), and sporangiotaxis in rows (character 197, states 1, 2).*
This character only applies to zosterophylls which have sporangia in vertical rows (Kenrick and Crane 1997).

199. MOD (KC5.19) **Orientation of lateral homosporous sporangia: Upright (0); Auricular (1).**
Rule 159*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and laterally attached sporangia (character 194, states 1-5).*
The auricular orientation seems to be a trait of a few zosterophylls and that the general condition among polysporangiates is upright (Kenrick and Crane 1997). As a result, ferns were scored upright. Bryophytes were scored not applicable. It is unclear whether this character applies beyond lycopsids and basal polysporangiates.

200. (S59) **Sorus: Absent (0); Present (1).**
Rule 160*. Contingent on the presence of a sporophyll (character 166, states 1, 2, 3).*
All non-ferns were scored absent.

201. (S60) **Sorus outline: Round (0); Elongate (1).**
Rule 161*. Contingent on the presence of sori (character 200, state 1).*
Taxa lacking sori were scored not applicable.

202. MOD (S61) **Sorus position: Abaxial, marginal to dorsal (0); Adaxial (1).**
Rule 162*. Contingent on the presence of sori (character 200, state 1).*
Taxa lacking sori were scored not applicable.

203. MOD (S62) **Sporangial maturation within sori: Simultaneous (0); Gradate (1); Mixed (2).**
Rule 163*. Contingent on the presence of sori (character 200, state 1).*
Taxa lacking sori were scored not applicable. Changed from the original character to exclude taxa lacking sori.

204. (S63) **Number of sporangia per sorus: Few, 1-12 (0); Many, more than 20 (1).**
Rule 164*. Contingent on the presence of sori (character 200, state 1).*
All taxa lacking sori were scored inapplicable.

205. (S65) **False indusium: Absent (0); Present (1).**
Rule 165*. Contingent on the presence of sori (character 200, state 1).*
All taxa lacking sori were scored not applicable. Original character altered so that it excluded taxa lacking sori.

206. (S66) **True indusium: Absent (0); Present (1).**
Rule 166*. Contingent on the presence of sori (character 200, state 1).*
All taxa lacking sori were scored not applicable.

207. (S67) **Attachment of true indusium relative to sori: Lateral (0); Basal (1); Central (2).**
Rule 167*. Contingent on the presence of a true indusium (character 206, state 1).*
All taxa lacking a true indusium were scored not applicable.

208. (S68) **Opening of true indusium: Introrse** (**0**)**; Extrorse** (**1**)**; Suprasoral** (**2**)**; Circumsoral** (**3**)**; None** (**4**)**.**
Rule 168*. Contingent on the presence of a true indusium (character 206, state 1).*
All taxa lacking a true indusium were scored not applicable.

209. MOD (S76) **Annulus: Absent** (**0**)**; Present – incomplete** (**1**)**; Present – complete** (**2**)**.**
Rule 169*. Contingent on the presence of homosporous sporogenesis (character 4, state 0) and the presence of sporangia (character 174, state 1).*
An incomplete annulus is a synapomorphy of leptosporangiate ferns. Mosses posses a complete annulus. All other taxa were scored absent.

210. (S77) **Annulus aspect on sporangium: Apical** (**0**)**; Lateral** (**1**)**; Oblique to transverse** (**2**)**; Vertical** (**3**)**.**
Rule 170*. Contingent on the presence of an incomplete annulus (character 209, state 1).*
All taxa lacking an annulus were scored not applicable.

211. (S78) **Annulus span across sporangium: Continuous bow** (**0**)**; Interrupted bow** (**1**)**; Restricted patch** (**2**)**.**
Rule 171*. Contingent on the presence of an incomplete annulus (character 209, state 1).*
All taxa lacking an annulus were scored not applicable.

212 – 226. ***Heterosporous sporangia***
212 – 220. ***Megasporangia***

212. (HB34) **Megasporangia/ovule -bearing structure: Pinnate** (**megasporangia, ovules or 'cupules' in two rows on a dorsiventral structure or pinnate with a three-dimensional fertile portion**) (**0**)**; Simple, paddle like** (**megasporangia or ovules not in two definite rows**) (**1**)**; Simple, stalk-like, with one megasporangium or ovule, or megasporangia/ovule sessile** (**2**)**.**
Rule 172*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Homosporous taxa were scored not applicable. Following Doyle (1996), angiosperms were scored as pinnate, though the Laurales and Chloranthaceae were scored unknown. Heterosporous lower plants were scored unknown.

213. (HB35) **Megasporangia/ovule attachment: On lateral appendage or sessile on lateral stem** (**0**)**; Terminal on stem** (**1**)**; Marginal** (**2**)**.**
Rule 173*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Homosporous taxa scored not applicable. Following the scoring of Doyle (1996) and Hilton and Bateman (2006), all angiosperms were scored as lateral.

214. (HB36) **Megasporangia/ovule position on supporting laminar structures: Apical or marginal** (**0**)**; Abaxial** (**1**)**; Adaxial** (**2**)**.**
Rule 174*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Following Doyle (1996), angiosperms were scored abaxial, though fossil taxa and Chloranthaceae were scored unknown.

215. (HB37) **Megasporangia/ovule orientation: Erect** (**0**)**; Inverted** (**1**)**.**
Rule 175*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Following the definition of Doyle (1996), angiosperms were scored erect.

216. (HB38) **Megasporangia/ovule enclosing structure: With no enclosing structure** (**0**)**; In a radial, lobed cupule** (**1**)**; In an anatropous cupule** (**2**)**; In an orthotropous cupule** (**3**)**; In a bipartite outer integument derived from two primordial** (**4**)**.**
Rule 176*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Angiosperms were scored based on DE115. Homosporous taxa scored not applicable.

217. (HB59) **Megasporangia dehiscence: Along one side** (**0**)**; Over the apex** (**1**)**; Indehiscent** (**2**)**.**
Rule 177*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Following HB, all angiosperms are scored as indehiscent. Heterosporous ferns (Salviniales) release spores via a gelatinous mass or tissue decay (Pryer 1999), and so were scored unknown. Lower heterosporous taxa were scored unknown.

218. (HB52) **Symmetry of the megasporangia-bearing or ovuliferous shoots: Radial** (**0**)**; Bilateral** (**1**)**.**
Rule 178*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Bilateral is the condition that unites certain conifers (Doyle 1996) and so all angiosperms were scored radial. Lower land plants were scored unknown.

219. (HB53) **Presence of appendages on megasporangia-bearing or ovuliferous shoots: With distinct appendages** (**0**)**; Without distinct appendages** (**1**)**.**
Rule 179*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
This is a character that distinguishes extant conifers (Doyle 1996) and so all angiosperms were scored (0). Lower heterosporous taxa were scored unknown.

220. (HB71) **Megasporangia/nucellus vascularization: Not vascularised** (**0**)**; Vascularised** (**1**)**.**
Rule 180*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
In the most extensively vascularised angiosperms, the vascularisation does not reach the nucellus (Gifford and Foster 1989; Endress 2011). Thus, all angiosperms were scored not vascularised. Lower land plants were scored unknown.

221 – 226. ***Microsporangia***

221. MOD (DE74) **Number of microsporangia per sporophyll / anther: Four** (**0**)**; Two** (**1**)**; Variable, two to twenty** (**2**)**; Twenty or more** (**3**)**; One** (**4**)**; Three** (**5**)**.**
Rule 181*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*

This character was expanded to accommodate the greater diversity of microsporangia found in other seed plants. Conifers and Gnetales were scored according to Biswas and Johnri (1997), with conifers possessing highly variable numbers, *Welwitschia* uniquely possessing 3 per microsporangiophore and Cycadales possessing much higher numbers. *Gnetum* is too highly modified to be interpreted (Doyle 1996). Fossil taxa were scored according to descriptions of Biswas and Johnri (1997) and Taylor *et al.* (2009). Pteridosperms generally possessed high numbers, with the exception of the Cordaitales and Caytoniales. *Archaeopteris* can have up to 40 sporangia per sporophyll, but the distribution of mega- and microsporophylls is unknown (Taylor *et al.* 2009). Similarly, stem and leaf taxa such were scored unknown. The heterosporous ferns have multiple sori per sporocarp (Kramer and Green 1990) and so were coded as many, despite *Marsilea* possessing few sporangia per sorus.

222. (HB44) **Position of microsporangia: Terminal (0); Abaxial (1); Adaxial (2); Lateral (3).**
Rule 182*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Following Doyle (1996), this character was altered to reflect only microsporangia and so all homosporous taxa were scored not applicable. Most angiosperms were scored lateral (Doyle 1996), with the exception of those scored adaxial by HB. Marsilea was scored abaxial (Schneider *et al.* 2009).

223. COMB (S80, D42) **Mechanism of sporangial dehiscence: Ectokinetic (0); Endokinetic (1); ~~Endothecial (2).~~**
Rule 183*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Heterosporous plants were coded based on microsporangia according to Doyle (1996) and Hilton and Bateman (2006), where all angiosperms were scored endothecial. Glossopterids were scored ectokinetic following Ryberg *et al.* (2012). Many mosses and lycophytes which were not scored in the original study were scored unknown.

224. MOD (DE75) **Orientation of anther dehiscence: Distinctly introrse (0) introrse to latrorse (1); Extrorse (2).**
Rule 184*. Contingent on the presence of a flower (character 227, state 1).*
This character was changed so that it refers only to angiosperm anthers.

225. (DE76) **Mode of anther dehiscence: Longitudinal slit (0); H-valvate (1); Valvate with upward-opening flaps (2).**
Rule 185*. Contingent on the presence of a flower (character 227, state 1).*
This character was changed so that it refers only to angiosperm anthers.

226. (HB 78) **Microspore/pollen cytokinesis: Successive (0); Simultaneous (1).**
Rule 186*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and the presence of sporangia (character 174, state 1).*
Scoring of angiosperms according to character 79 of DE. *Isoetes* and *Selaginella* were scored successive (Heng-Chang *et al.* 2007). All homosporous taxa and green algae scored not applicable. Fossil taxa scored as unknown. The heterosporous ferns, *Salvinia* and *Azolla* were scored as unknown.

227 – 278. *Flowers and Inflorescences*
On flowers: For all floral characters, the scoring of Doyle and Endress (2010) was used, with the exception of Oryza sativa *and* Glycine max *which were included from the study of (Mishler et al. 1994). Additionally,* Sinocarpus*, a fossil angiosperm, was included. All scoring of this taxon was based on Hilton and Bateman (2006) and Leng and Friis (2006).*

227. (M110) **Flowers: Absent** (**0**)**; Present** (**1**)**.**
Rule 187*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1).*
As a synapomorphy of angiosperms, all non-angiosperm taxa were scored absent (Doyle 2013).

228. (DE42) **Inflorescence: Solitary flower, occasional lateral flowers** (**0**)**; Botryoid, thyrsoid or panicle, monotelic** (**1**)**; Raceme, spike or thyrse, polytelic** (**2**)**.**
Rule 188*. Contingent on the presence of a flower (character 227, state 1).*
All taxa lacking a flower were scored not applicable. The spikelets of *Oryza* were here treated as a type of panicle, whereas the inflorescence of *Glycine* develops as a raceme (Singh 2010).

229. (DE43) **Inflorescence partial units: Single flowers** (**0**)**; Cymes** (**1**)**.**
Rule 189*. Contingent on the presence of inflorescence (character 228, states 0 - 2).*
All taxa lacking a flower or inflorescence scored not applicable.

230. (DE44) **Inflorescence or partial inflorescence: Not modified** (**0**)**; Modified into a globular head** (**1**)**.**
Rule 190*. Contingent on the presence of inflorescence (character 228, states 0 - 2).*
All taxa lacking a flower or inflorescence scored not applicable. Scored as a feature of *Platanus* and some fossil platanoids (Doyle and Endress 2010), *Oryza* and *Glycine* were scored absent.

231. (DE45) **Pedicel: Present in some or all flowers** (**0**)**; Absent or reduced** (**1**)**.**
Rule 191*. Contingent on the presence of flowers (character 227, state 1).*
All taxa lacking a flower or inflorescence scored not applicable. Pedicels are recorded as present in *Oryza* and *Glycine* (Kuang *et al.* 1991; Terrell *et al.* 2001).

232. (DE46) **Floral subtending bracts: Present in all flowers** (**0**)**; Present in female, absent in male** (**1**)**; Absent in all flowers** (**2**)**.**
Rule 192*. Contingent on the presence of flowers (character 227, state 1).*
In scoring *Oryza,* the lemma and pelea were considered homologous to the sepal and bract structures (Simpson 2011). *Glycine* also possess bracts (Crozier and Thomas 1993).

233. (DE47) **Sex of flowers: Bisexual** (**0**)**; Unisexual** (**1**)**.**
Rule 193*. Contingent on the presence of flowers (character 227, state 1).*
All taxa lacking a flower or inflorescence scored not applicable. The flowers of *Oryza* and *Glycine* were both scored bisexual.

234. (DE48) **Floral base: Hypanthium absent, superior ovary** (**0**)**; Hypanthium present, superior ovary** (**1**)**; Partially or completely inferior ovary** (**2**)**.**
Rule 194*. Contingent on the presence of flowers (character 227, state 1).*
Both *Oryza* and *Glycine* possess a superior ovary, with *Glycine* possessing a hypanthium, but *Oryza* lacking one (Simpson 2011).

235. (DE49) **Floral receptacle: Short** (**0**)**; Elongate** (**1**)**.**
Rule 195*. Contingent on the presence of flowers (character 227, state 1).*
Following the scoring of (Doyle and Endress 2010), an elongate receptacle seems limited to several basal lineages. Consequently, both *Oryza* and *Glycine* were scored short.

236. (DE50) **Pits in receptacle bearing individual carpels: Absent** (**0**)**; Present** (**1**)**.**
Rule 196*. Contingent on the presence of flowers (character 227, state 1).*
Following Doyle and Endress (2010), all angiosperms except *Nelumbo* were scored absent.

237. (DE51) **Cortical vascular system: Absent or supplying the perianth only** (**0**)**; Supplying the androecium** (**1**)**; Supplying androecium and gynoecium** (**2**)**.**
Rule 197*. Contingent on the presence of flowers (character 227, state 1).*
Glycine and *Oryza* were scored unknown.

238. (DE52) **Floral apex: Used up after the production of the flower** (**0**)**; Protruding into mature flower** (**1**)**.**
Rule 198*. Contingent on the presence of flowers (character 227, state 1).*
Unicarpellate taxa were scored unknown. Following Doyle and Endress (2010), this character seems a feature only of some basal lineages of angiosperms. Consequently, *Glycine* and *Oryza* were scored (0).

239. (DE53) **Perianth: Present** (**0**)**; Absent** (**1**)**.**
Rule 199*. Contingent on the presence of flowers (character 227, state 1).*
The presence of a perianth among Poaceae is not clear, since the floral structure is so highly derived. If the lemma and pelea are interpreted as homologous to sepals and bracts, and the lodicule as homologous to petals, then it is present (Simpson 2011). In the case of *Oryza*, lodicules are present and so the perianth was scored as present.

240. (DE54) **Perianth phyllotaxis: Spiral** (**0**)**; Whorled** (**1**)**.**
Rule 200*. Contingent on the presence of a perianth (character 239, state 0).*
Taxa lacking a perianth were scored not applicable. *Oryza* was scored as whorled (Yamaguchi *et al.* 2006) as was *Glycine* (Crozier and Thomas 1993).

241. (DE55) **Perianth whorls or series: One** (**0**)**; Two** (**1**) **More than two** (**2**)**.**
Rule 201*. Contingent on the presence of a perianth (character 239, state 0).*
Taxa without a perianth were scored not applicable. *Oryza* was scored as having a single whorl, with the lodicules the only whorl.

242. (DE56) **Perianth merism: Trimerous** (**0**)**; Dimerous** (**1**)**; Polymerous** (**2**)**.**
Rule 202*. Contingent on the presence of a perianth (character 239, state 0).*
Glycine was scored pentamerous (Singh 2010). The irregular nature of the *Oryza* perianth meant that it was scored unknown.

243. (DE57) **Tepal differentiation: All sepaloid** (**0**)**; Outer sepaloid, inner petaloid** (**1**)**; All petaloid** (**2**)**.**
Rule 203*. Contingent on the presence of flowers (character 227, state 1).*
Glycine was scored as having differentiated tepals (Crozier and Thomas 1993). *Oryza* was scored as unknown, possessing neither petals nor sepals.

244. (DE58) **Petals: Absent** (**0**)**; Present** (**1**)**.**
Rule 204*. Contingent on the presence of flowers (character 227, state 1).*
Glycine was scored as present, while *Oryza* was scored as unknown.

245. (DE59) **Nectaries on inner perianth parts: Absent** (**0**)**; Present** (**1**)**.**
Rule 205*. Contingent on the presence of a perianth (character 239, state 0).*
Based on the description of Poaceae (Simpson 2011), *Oryza* was scored absent. *Glycine* does possess nectaries but they are positioned between the inner two whorls (Crozier and Thomas 1993).

246. (DE60) **Outermost perianth parts: Free** (**0**)**; Fused, at least basally** (**1**)**.**
Rule 206*. Contingent on the presence of a perianth (character 239, state 0).*
Oryza was scored unknown due to the irregular nature of the perianth. *Glycine* was scored as fused, as within Fabaceae the outermost perianth is fused into a calyx tube (Simpson 2011).

247. (DE61) **Calyptra derived from bracteate organs: Absent** (**0**)**; Present** (**1**)**.**
Rule 207*. Contingent on the presence of a perianth (character 239, state 0).*
Following (Doyle and Endress 2010), this is a feature of Magnoliales and so all other taxa were scored absent.

248. MOD (HB43, HB45, HB49) **Stamens: Absent** (**0**)**; Present** (**1**)**.**
Rule 208*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1).*
All taxa lacking a flower were scored absent.

249. (HB43) **Stamen form: Laminar** (**0**)**; Differentiated into anther and filament** (**1**)**.**
Rule 209*. Contingent on the presence of stamens (character 248, state 1).*
Taxa lacking stamens were scored not applicable. Taxa were scored according to Endress and Hufford (1989).
Taxa lacking well differentiated filaments but not classed as laminar were scored unknown.

250. (DE62) **Stamen number: More than one** (**0**)**; One** (**1**)**.**
Rule 210*. Contingent on the presence of stamens (character 248, state 1).*
All taxa lacking stamen were scored unknown.

251. (DE63) **Androecium phyllotaxis: Spiral** (**0**)**; Whorled** (**1**)**.**
Rule 211*. Contingent on the presence of stamens (character 248, state 1).*
Both *Glycine* and *Oryza* were scored as whorled (Crozier and Thomas 1993; Zhang and Wilson 2009).

252. (DE64) **Androecium merism: Trimerous** (**0**)**; Dimerous** (**1**)**; Polymerous** (**2**)**.**
Rule 212*. Contingent on the presence of stamens (character 248, state 1).*
Glycine is pentamerous (Skorupska *et al.* 1993) while *Oryza* has a trimerous structure (Dahlgren *et al.* 1985).

253. (DE65) **Number of stamen whorls: One** (**0**)**; Two** (**1**)**; More than two** (**2**)**.**
Rule 213*. Contingent on the presence of stamens (character 248, state 1).*
Both *Glycine* and *Oryza* have two whorls of stamen (Singh 2010; Verma 2010).

254. (DE66) **Stamen positions: Single** (**0**)**; Double** (**1**)**.**
Rule 214*. Contingent on the presence of stamens (character 248, state 1).*
Based on the number of whorls (two) and number of stamen (10, six), it was determined that both *Glycine* and *Oryza* respectively had single positions.

255. (DE67) **Stamen fusion: Free** (**0**)**; Connate** (**1**)**.**
Rule 215*. Contingent on the presence of stamens (character 248, state 1).*
Nine of the ten filaments in *Glycine* are basally fused (Skorupska *et al.* 1993) so was scored connate. *Oryza* was scored free based on the generalised spikelet of Poaceae (Simpson 2011).

256. COMB (HB50, DE68) **Inner staminodes: Absent** (**0**)**; Present** (**1**)**.**
Rule 216*. Contingent on the presence of stamens (character 248, state 1).*
No record of staminodes was in either *Glycine* or *Oryza* and so both were scored absent.

257. (DE69) **Glandular food bodies on stamen or staminodes: Absent** (**0**)**; Present** (**1**)**.**
Rule 217*. Contingent on the presence of stamens (character 248, state 1).*
Based on the scoring by Doyle and Endress (2010) of monocots and eudicots, *Oryza* and *Glycine* were both scored absent.

258. (DE70) **Stamen base: Short** (**2/3 or less the length of the anther**) (**0**)**; Long and wide** (**More than ½ width of the anther**) (**1**)**; Long and narrow** (**2**)**.**
Rule 218*. Contingent on the presence of stamens (character 248, state 1).*
Both *Oryza* and *Glycine* were scored long and narrow (Crozier and Thomas 1993; Simpson 2011).

259. (DE71) **Paired basal stamen glands: Absent** (**0**)**; Present** (**1**)**.**
Rule 219*. Contingent on the presence of stamens (character 248, state 1).*
Based on the scoring of Doyle and Endress (2010), this is a trait confined to Laurales, so *Oryza* and *Glycine* were scored absent.

260. (DE72) **Connective apex: Extended** (**0**)**; Truncated/smoothly rounded** (**1**)**; Peltate** (**2**)**.**
Rule 220*. Contingent on the presence of stamens (character 248, state 1).*
Information for *Oryza* and *Glycine* could not be found and both were scored unknown.

261. (DE73) **Pollen sacs: Protruding** (**0**)**; Embedded** (**1**)**.**
Rule 221*. Contingent on the presence of stamens (character 248, state 1).*
Information for *Oryza* and *Glycine* could not be found and both were scored unknown.

262. (HB39) **Closed carpel with stigmatic germination: Absent** (**0**)**; Present** (**1**)**.**
Rule 222*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1).*
All non-angiosperm taxa were scored absent.

263. (DE96) **Carpel number: One** (**0**)**; 2-5 in one whorl** (**1**)**; More than 5 in one whorl** (**2**)**; More than one whorl** (**3**)**.**
Rule 223*. Contingent on the presence of closed carpels (character 262, state 1).*
Based on general descriptions of Poaceae and Fabaceae (Simpson 2011), *Glycine* is unicarpellate, and *Oryza* has 2-3 carpels in a single whorl.

264. (DE97) **Carpel form: Ascidiate up to stigma** (**0**)**; Intermediate, both plicate and ascidiate zones below the stigma, ovules on ascidiate zone** (**1**)**; Completely plicate or with ovules on the plicate zone** (**2**)**.**
Rule 224*. Contingent on the presence of closed carpels (character 262, state 1).*
Information for *Oryza* and *Glycine* could not be found and both were scored unknown.

265. (DE98) **Postgenital sealing of carpel: None (0); Partial (1); Complete (2).**
Rule 225*. Contingent on the presence of closed carpels (character 262, state 1).*
Information for *Oryza* and *Glycine* could not be found and both were scored unknown.

266. (DE99) **Secretion in area of carpel sealing: Present (0); Absent (1).**
Rule 226*. Contingent on the presence of closed carpels (character 262, state 1).*
Information for *Oryza* and *Glycine* could not be found and both were scored unknown.

267. (DE106) **Carpel fusion: Apocarpous (0); parasyncarpous (1); Eusyncarpous (2).**
Rule 227*. Contingent on the presence of closed carpels (character 262, state 1).*
Unicarpellate taxa were scored unknown. *Oryza* was scored eusyncarpous (Simpson 2011).

268. (DE107) **Oil cells in carpels: Absent or internal (0); Intrusive (1).**
Rule 228*. Contingent on the presence of closed carpels (character 262, state 1).*
Along with the other eudicots and monocots, *Oryza* and *Glycine* were scored unknown.

269. (DE108) **Long unicellular hairs on/between carpels: Absent (0); Present (1).**
Rule 229*. Contingent on the presence of closed carpels (character 262, state 1).*
There was no record of unicellular hairs for *Oryza* in the literature and so, as with other monocots, it was scored absent. Hairs are present on the carpel of *Glycine* (Healy *et al.* 2005).

270. (DE109) **Short curved, appressed, unlignified hairs with up to two short basal cells and one long apical cell on carpels: Absent (0); Present (1).**
Rule 230*. Contingent on the presence of closed carpels (character 262, state 1).*
As no reference was made to these structure in either *Oryza* or *Glycine* (Healy *et al.* 2005), both were scored absent.

271. (DE110) **Nectary on dorsal or lateral side of carpel or pistillode: Absent (0); Present (1).**
Rule 231*. Contingent on the presence of closed carpels (character 262, state 1).*
Oryza lacks nectaries (Simpson 2011), while those of *Glycine* are not located on the carpel (Horner *et al.* 2003).

272. (DE111) **Septal nectaries or potentially homologous intercarpellary nectaries: Absent (0); Present (1).**
Rule 232*. Contingent on the presence of closed carpels (character 262, state 1).*
The position of the nectaries of *Glycine* between the carpel and the lateral stamens means it was scored absent (Horner *et al.* 2003), as was *Oryza*.

273. (DE112) **Number of ovules per carpel: One (0); Two, or varying between one and two (1); More than two (2).**
Rule 233*. Contingent on the presence of closed carpels (character 262, state 1).*
Glycine possesses up to 4 ovules per carpel (Kennell and Horner 1985), the fossil *Sinocarpus* has up to 20 (Leng and Friis 2006) and *Oryza* has a single ovule per carpel (Yamaguchi *et al.* 2006).

274. (DE101) **Style: Absent, carpel sessile (0); Present (1).**
Rule 234*. Contingent on the presence of closed carpels (character 262, state 1).*
Both *Glycine* and *Oryza* were scored present (Tilton *et al.* 1984); Yamaguchi et al., 1993).

275. (DE102) **Stigma: Extended, half or more of the style-stigma zone (0); Restricted (1).**
Rule 235*. Contingent on the presence of closed carpels (character 262, state 1).*
Information for *Oryza* and *Glycine* could not be found and both were scored unknown.

276. (DE103) **Multicellular stigmatic protuberances or undulations: Absent (0); Present (1).**
Rule 236*. Contingent on the presence of closed carpels (character 262, state 1).*
The absence of these structures was inferred from descriptions of stigma in both *Glycine* (Tilton *et al.* 1984) and *Oryza* (Ciampolini *et al.* 2001).

277. (DE104) **Stigmatic papillae: Absent (0); Unicellular, or with a single emergent cell and more than one basal cell (1); Multicellular with the emergent portion consisting of more than one cell (2).**
Rule 237*. Contingent on the presence of closed carpels (character 262, state 1).*
Oryza possesses multicellular papillae (Ciampolini *et al.* 2001) while *Glycine* has unicellular (Tilton *et al.* 1984).

278. (DE105) **Extragynoecial compitum: Absent (0); Present (1).**
Rule 238*. Contingent on the presence of closed carpels (character 262, state 1).*
As there were no references to a compitum in either *Glycine* or *Oryza*, both were scored absent.

279 – 299. ***Roots, mycorrhiza and prokaryotes***
On roots: Though the rooting structures of lycophytes, ferns and seed plants are not homologous (Rothwell and Erwin 1985), they all possess a functionally and morphologically similar structure which is defined as a root. For this reason, all are scored present in character 280. However, for subsequent characters that refer to specific features of either lycophyte (281-282) or seed plant roots (283), each is respectively scored not applicable, due to the fundamental differences.

279. (S36) **Roots: Absent (0); Present (1).**
Early tracheophytes scored according to Kenrick and Crane (1997) and Gensel and Edwards (2001). *Psilophyton* possesses structures reminiscent of roots (Stewart and Rothwell 2010), but was scored unknown. Zosterophylls were scored as present, due to the likely synapomorphy of root tufts (Hao *et al.* 2010). With the exception of *Ceratophyllum* (Gluck 1906), angiosperms were scored as present (Seago and Fernando 2013). *Lyginopteris*, *Heterangium, Medullosa*, Cordaitales, Glossopterids were scored as present according to descriptions of Taylor *et al.* (2009). Similarly, Bennettitales, *Callistophyton* and Cheirolepidiaceae were scored present (Rothwell 1975; Axsmith and Jacobs 2005; Strullu-Derrien *et al.* 2012). Taxa without a clear description were scored unknown. Many stem taxa e.g. *Pentoxylon* or large conifers (*Emporia/Thucydia*) would likely have had roots, and were scored present. All algae and non-vascular plants scored absent.

280. (S37) **Rhizomorphs/rhizophores: Absent (0); Present (1).**
This structure is confined to rhizomorphic lycopsids (Schneider *et al.* 2009), and is reported in *Paralycopodites* (DiMichele and Bateman 1996). All other taxa with preserved roots scored absent.

281. (KC6.32) **Root stele symmetry: Radial (0); Bilateral (1).**
Rule 239*. Contingent on the presence of roots (character 279, state 1).*
All ferns and seed plants are scored not applicable due to the fundamental differences in rooting structures between the roots of lycophytes and other lineages (Rothwell and Erwin 1985).

282. (KC6.33) **Root xylem shape: More or less circular** (**0**)**; Crescent shaped** (**1**)**.**
Rule 240*. Contingent on the presence of roots (character 279, state 1).*
Progymnosperms and seed plants scored not applicable.

283. (KC6.34) **Cortical roots: Absent** (**0**)**; Present** (**1**)**.**
Rule 241*. Contingent on the presence of roots (character 279, state 1).*
These structures are described as unique to *Huperzia* (Kenrick and Crane 1997) and *Phlegmariurus* and so were scored absent for all other taxa.

284. (HB3) **Radicle: Persistent** (**0**)**; Replaced by an adventitious root** (**1**)**.**
Rule 242*. Contingent on the presence of roots (character 279, state 1).*
As the roots of lycophytes and ferns do not originate from an embryonic radicle, they were scored not applicable. Monocots and Nymphaeales tend to lose their radicle, with other taxa scored as persistent (Seago and Fernando 2013). All fossil taxa scored as unknown.

285. (S38) **Root origin: Primarily allorhizic** (**0**)**; Homorhizic or secondarily allorhizic** (**1**)**.**
Rule 243*. Contingent on the presence of roots (character 279, state 1).*
Non-seed plants possess a unipolar embryo and so were scored primarily allorhizic (Schneider *et al.* 2009). As seed plants are characterised by a bipolar embryo, all extant seed plants were scored as homorhizic or secondarily allorhizic (Gifford and Foster 1989) (Schneider *et al.* 2009). All fossil taxa were scored as unknown.

286. (S39) **Root branching: Dichopodial/exogenous** (**0**)**; Monopodial/endogenous** (**1**)**.**
Rule 244*. Contingent on the presence of roots (character 279, state 1).*
Dichopodial branching is a unique feature of the lycophytes, while euphyllophytes show monopodial branching (Gifford and Foster 1989). All fossil taxa were scored unknown.

287. (S40) **Lateral root origin in endodermis: Absent** (**0**)**; Present** (**1**)**.**
Rule 245*. Contingent on the presence of monopodial branching in roots (character 286, state 1).*
This character is only applicable to taxa with monopodial branching and represents a distinction between ferns and seed plants (Schneider *et al.* 2009). All fossil taxa scored unknown.

288. (S41) **Root hairs: Present** (**0**)**; Absent** (**1**)**.**
Rule 246*. Contingent on the presence of roots (character 279, state 1).*
Though rare and uncommon in certain taxa such as *Isoetes* and *Marsilea*, root hairs are only absent in Ophioglossaceae (Schneider *et al.* 2009). Scored unknown in fossil taxa.

289. (S42) **Root hair structure: Non-septate** (**0**)**; Septate** (**1**)**.**
Rule 247*. Contingent on the presence of root hairs (character 288, state 0).*
Septate root hairs are caused by the formation of septa within the unicellular root hair, giving the impression of multicellularity, and only occur within the Marattiaceae (Schneider *et al.* 2009). Fossil taxa scored unknown.

290. (S43) **Rhizodermis cells: Undifferentiated** (**0**)**; Differentiated into long and short cells** (**1**)**.**
Rule 248*. Contingent on the presence of roots (character 279, state 1).*

Undifferentiated rhizodermis is the common state in most angiosperms, with differentiated found only in some monocots and Nymphaceae (Leavitt 1904). Information is currently lacking for gymnosperms and so they were left unknown.

291. (S45) **Number of protoxylem poles in root: Variable, ranging from monarch to 18-arch** (**0**)**; Variable, ranging from monarch to hexarch** (**1**)**; Usually diarch, rarely triarch** (**2**)**; Predominantly monarch** (**3**)**.**
Rule 249*. Contingent on the presence of roots (character 279, state 1).*
Monocots and Nymphaceae have predominantly polyarch roots (Seago and Fernando 2013), whereas eudicots and Magnoliids tend to have diarch to hexarch roots (Esau 1977; Seago and Fernando 2013). A notable exception is *Nelumbo* which has a highly polyarch root (Esau 1977). *Amborella* features a diarch root, though other basal taxa such as Cabombaceae and Hydatellaceae have a strictly monarch root (Seago 2002). Remaining angiosperms lack information and were scored unknown. Of fossil taxa, *Lyginopteris* and *Heterangium* are triarch to polyarch, while *Medullosa* and *Cordaitales* are triarch to pentarch and *Callistophyton* is diarch (Taylor *et al.* 2009). Based on *Vertebraria,* Glossopterids are diarch to 7-arch (Decombeix *et al.* 2009) and Bennettitales are diarch to pentarch (Strullu-Derrien *et al.* 2012). As a rule, gymnosperms are diarch (Stevenson 2012) and lycophytes have a type of diarch (Carlquist 2001). Remaining taxa lacking information were scored unknown.

292. (S46) **Aerenchyma in root cortex: Absent** (**0**)**; Present, septate cells not differentiated** (**1**)**; Present, septate cell differentiated** (**2**)**.**
Rule 250*. Contingent on the presence of roots (character 279, state 1).*
Septate aerenchyma is unique to Marsileaceae (Schneider *et al.* 2009). Though a thorough survey of aerenchyma in seed plants is lacking, aerenchyma is found in some wetland and aquatic species as outlined by (Seago *et al.* 2005), though absent in Magnoliids and Amborella (Seago and Fernando 2013). Remaining taxa were scored unknown.

293. (S44) **Root pith: Absent** (**0**)**; Present** (**1**)**.**
Rule 251*. Contingent on the presence of roots (character 279, state 1).*
In leptosporangiate ferns, the vascular tissue forms a compact circle with a central pith in the center of the root cross-section, whereas non-vascular tissue is central in Equisetopsida, Marattiopsida and Ophioglossopsida. A pith is also present in most gymnosperms, such as Cycas, and some angiosperms (monocots), but absent in eudicots (von Guttenberg 1968a, 1968b; Ogura 1972; Esau 1977; Fahn 1990; Stevenson 1990; Schneider 1996; Norstog and Nichols 1997).

294. NEW Obligate mycorrhizas in sporophytes: Absent (**0**)**; Present** (**1**)**.**
Rule 252*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*

295. NEW Fungal endophyte: septate (**0**)**; aseptate** (**1**)**.**
Rule 253*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
This character will be not applicable to bryophytes.

296. NEW Septate fungal endophyte: Ascomycota (**0**)**; Basidiomycota** (**1**)**.**
Rule 254*. Contingent on the presence of septate fungal endophytes (character 295, state 0).*

297. NEW Aseptate fungal endophyte: Mucoromycotina (**0**)**; Glomeromycota** (**1**)**; Mucoromycotina** + **Glomeromycota** (**dual partnership**) (**2**)**.**
Rule 255*. Contingent on the presence of aseptate fungal endophytes (character 295, state 1).*

298. Mycoheterotrophy: Present (0), Absent (1).
Rule 256*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*
Note that for some taxa e.g. *Lycopodiella* we will need to score multiple states.

299. Prokaryote endophytes: Cyanobacteria (0), rhizobia (1), actinobacteria (2)
Rule 257*. Contingent on the presence of a multicellular sporophyte (character 43, state 1).*

IV. Gametophyte (300 - 357)

300 – 326. ***General gametophyte features and development***

300. NEW Multicellular gametophyte: Absent (0); Present (1).

301. (M18) **Gamete production: Holocarpic (0); Heterocarpic (1).**
All charophytes and embryophytes were scored heterocarpic.

302. (KC3.16) **Gametophyte form: Thalloid (0); Axial (1); Highly reduced (2).**
Rule 258*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Ferns scored using S99. All seed plants scored highly reduced (Doyle 2013). Fossil polysporangiates lacking a well-preserved gametophyte and basal tracheophytes scored unknown.

303. COMB (S50, PSS81) **Gametophyte apical meristem structure: Unicellular or up to 4 initial cells (0); Complex and more than 4 initial cells (1).**
Rule 259*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

304. NEW Gametophytic extra-mural material: Absent (0); Present - differentiated (1); Present - undifferentiated (2).
Rule 260*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

305. NEW Schizogenous intercellular spaces within gametophyte: Absent (0); Present (1).
Rule 261*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

306. NEW Gametophytic water-conducting cells: Absent (0); Present (1).
Rule 262*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

307. NEW Thickness of primary water-conducting cell walls: Thin (0); Thick (1).
Rule 263*. Contingent on the presence of gametophytic water-conducting cells (character 306, state 1).*

308. NEW Perforate primary walls of water-conducting cells: Absent (0); Present (1).
Rule 264*. Contingent on the presence of gametophytic water-conducting cells (character 306, state 1).*

309. NEW Pitting size in the primary walls of perforate water-conducting cells: Small (0); Large (1).
Rule 265*. Contingent on the presence of gametophytic water-conducting cells (character 306, state 1).*

310. NEW Gametophytic food-conducting cells: Absent (0); Present (1).
Rule 266*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

311. NEW Microtubular cytoskeleton of gametophytic food-conducting cells: Actin-dominant (0), Tubulin-dominant (1).
Rule 267*. Contingent on the presence of gametophytic food-conducting cells (character 310, state 1).*
This will distinguish between the FCCs in bryophyte gametophyte (1) and FCCs in pteridophyte gametophyte (2) – only in *Tmesipteris* and *Psilotum*.

312. COMB (KC3.17, M98) **Gametophyte leaves: Absent (0); Present (1).**
Rule 268*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
All extant vascular plants and those with clearly preserved gametophytes were scored absent, including fossil seed plants.

313. (KC3.25) **Protonema type: Nonfilamentous or undifferentiated (0); Elaborate, persistent, filamentous (1).**
Rule 269*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Most ferns germinate to a filamentous protonema (Wada 2008), with the exception of the Osmundaceae and Marattiaceae (Raghavan 1989). Hornworts lack a protonema, and the condition in most liverworts is globose (Glime 2007).

314. (KC3.26, M87) **Oil bodies: Absent (0); Present (1).**
Rule 270*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
The oil bodies of liverworts seem to be fundamentally distinct from every other type of lipid body found in land plants (Kenrick and Crane 1997).

315. NEW Position of oil bodies: Every cell (0); Restricted to idioblasts (1).
Rule 271*. Contingent on the presence of oil bodies (character 314, state 1).*

316. (KC3.28) **Pseudopodium: Absent (0); Present (1).**
Rule 272*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Both *Andreaea* and *Sphagnum* are unique amongst land plants in possessing a pseudopodium (Kenrick and Crane 1997).

317. MOD (KC6.39) **Gametophyte habit: Superficial (0); Subterranean (1); Aqueous (2).**
Rule 273*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
A unique type of subterranean, holosaprotrophic gametophyte occurs within the Lycopodiaceae (Kenrick and Crane 1997), and so all other taxa were scored superficial.

318. (HB89) **Sterile cell: Colinear with other microgametophyte cells (0); Ring-shaped (1).**

Rule 274*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and a multicellular gametophyte (character 300, state 1).*
All homosporous taxa were scored not applicable. Following Hilton and Bateman (2006), all angiosperms and most fossils were scored unknown.

319. (HB91) **Megagametophyte spores: Monosporic** (**0**)**; Tetrasporic** (**1**)**.**
Rule 275*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and a multicellular gametophyte (character 300, state 1).*
All homosporous taxa were scored unknown. All angiosperms included in the present study, with the exception of Piperaceae, were scored monosporic (Friedman and Williams 2004).

320. (HB93) **Megagametophyte cellularization: Enclosing single nuclei, resulting in uninucleate cells** (**0**)**; Enclosing several nuclei, resulting in polyploid cells** (**1**)**.**
Rule 276*. Contingent on the presence of heterosporous sporogenesis (character 4, state 1) and a multicellular gametophyte (character 300, state 1).*
Following the scoring of Hilton and Bateman (2006), all angiosperms were scored as uninucleate.
321. (DE136) **Female gametophyte: Eight to nine nuclei** (**0**)**; Four nuclei** (**1**)**.**
Rule 277*. Contingent on the presence of flowers (character 227, state 1).*
This character was applied only to angiosperms as a means to differentiate embryo types, particularly those among basal angiosperms. The diversity of megagametophytes among other land plants was captured in character S124.

322. (S100) **Gametophyte with *Gleichenia* type club-shaped hairs: Absent** (**0**)**; Present** (**1**)**.**
Rule 278*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
This type of development is known only from the gametophytes of the Gleicheniaceae (Schneider *et al.* 2009).

323. (S101) **Gametophyte with bristle- to scale-like hairs: Absent** (**0**)**; Present** (**1**)**.**
Rule 279*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
This type of development is known only in the Loxomataceae and Cyatheaceae (Schneider *et al.* 2009).

324. (KC6.37) **Gamatophyte development: Exosporic** (**0**)**; Endosporic** (**1**)**.**
Rule 280*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Due to the nature of the seed, all seed plants were scored endosporic. Heterosporous ferns were also scored endosporic (Wolniak *et al.* 2011). All homosporous plants were scored exosporic.

325. (S104) **Gametangia distribution: Widely distributed, non-terminal** (**0**)**; Terminal** (**1**)**.**
Rule 281*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Terminal gametangia are a key differentiator between lycopsids and euphyllophytes (Schneider *et al.* 2009). Consequently, all euphyllophytes were scored as possessing non-terminal gametangia. Among bryophytes, the Marchantiopsida and Haplomitriopsida have broadly distributed gametangia, while the Jungermanniopsida have terminal androecium (Crandall-Stotler *et al.* 2009). The gametangia of mosses are always terminal (Goffinet and Shaw 2009).

326. (S124) **Megagametophyte form: Large, completely cellular, with cellular archegonia** (**0**)**; Large, apical portion with egg free-nuclear** (**1**)**; Multi-nucleate, central portion free-nuclear, egg cellular but without neck cells** (**2**)**.**

Rule 282. *Contingent on the presence of a multicellular gametophyte (character 300, state 1) and heterosporous sporogenesis (character 4, state 1).*
All angiosperms were scored (2), including fossil taxa. As the condition varies across gymnosperms and seed plants, fossils were scored unknown.

327 – 330. ***Mucilage cells***

327. NEW Mucilage cells: Present (0); Absent (1).
Rule 283. *Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

328. NEW Mucilage cell shape: Flask shaped (0); Rounded (1).
Rule 284. *Contingent on the presence of mucilage cells (character 327, state 0).*

329. NEW Mucilage cells number: 2 (0); More than 2 (1).
Rule 285. *Contingent on the presence of mucilage cells (character 327, state 0).*
330. (KC3.27) **Mucilage clefts: Absent (0); Present (1).**
Rule 286. *Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
The Anthocerotopsida and *Blasia* are characterised by mucilage cavities (Kenrick and Crane 1997). All other taxa were scored absent.

331 – 334. ***Archegonia***

331. COMB (KC3.18, M81) **Archegonium: Absent (0); Present (1).**
Rule 287. *Contingent on the presence of alternation of generations (character 3, states 2, 3).*
An archegonium in some form is a feature of all embryophytes groups (Kenrick and Crane 1997). Taxa lacking a clearly preserved gametophyte were scored unknown.

332. (KC3.21) **Archegonium position: Superficial (0); Sunken (1).**
Rule 288. *Contingent on the presence of archegonium (character 331, state 1).*
The archegonia of all liverworts and mosses are superficial, all other taxa scored sunken (Kenrick and Crane 1997).

333. (S122) **Number of archegonium neck cell tiers: More than six (0); One to five, rarely six (1).**
Rule 289. *Contingent on the presence of archegonium (character 331, state 1).*
All angiosperms were scored not applicable, as the neck cells are represented only by the synergids. Gymnosperms were scored as having between one and six (Biswas and Johnri 1997). Liverworts belonging to Marchantiopsida and Jungermanniopsida possess 6 and 5 rows respectively (Crandall-Stotler *et al.* 2009). Hornworts uniformly possess 6 rows (Renzaglia *et al.* 2009). It is reported that the general condition in mosses is 10 layers or more (Rashid 1998), with the exception of *Sphagnum*, *Oedipodium* and *Polytrichum*. Fossil taxa scored unknown, except for *Horneophyton*, for which many cell layers are reported (Remy *et al.* 1993).

334. MOD (S123) **Neck canal cell: Multinucleate (0); Binucleate (1); Mononucleate (2).**
Rule 290. *Contingent on the presence of archegonium (character 331, state 1).*
In bryophytes, the neck canal cells are mononucleate. Most ferns possess a single binucleate neck canal cell, although some homosporous ferns can possess neck canal cells that are multinucleate (Raghavan, 2006).

335 – 348. *Antheridia*

335. (GRD1) **Multicellular antheridia: Absent (0); Present (1).**
Rule 291. *Contingent on the presence of alternation of generations (character 3, states 2, 3).*
After Garbary *et al.* (1993): based on the development of the male structure of *Chara* (Pickett-Heaps 1968), the male reproductive organs of *Chara* and *Nitella* are considered to be true antheridia. The distinct jacket cells in *Selaginella* and *Marsilea* are regarded as single, reduced antheridia (Bierhorst 1971).

336. (GRD2, RG2) **Apical cell in antheridial ontogeny: Absent (0); Present (1).**
Rule 292. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
After Garbary *et al.* (1993): An apical cell is present in the developing antheridia of *Takakia,* as it is in other mosses, including *Sphagnum* (Smith 1955; Schofield 1985). Not present in the green algae, liverworts, hornworts or higher land plant taxa.

337. (GRD3) **Division patterns in the young antheridia: Four-celled pattern (0); Two-celled pattern (1).**
Rule 293. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
After Garbary *et al.* (1993): "This refers to the series of cell divisions that delimit the primary androgones in cross-sectional view of the antheridial body initial. In the two-celled pattern, two androgones originate with the peripheral segmentation of four jacket cells. In the four-celled pattern, quadrants of cells are delimited and two jacket cells then form in each quadrant; the four primary androgones are therefore surrounded by eight jacket cells."

338. (GRD4) **Arrangement of spermatogenous cells: Single (0); Filaments (1); Blocks (2); Random (3).**
Rule 294. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
After Garbary *et al.* (1993): "Blocks of spermatogenous cells that represent the initial cell division in organogenesis are characteristic of bryophyte antheridia and are most prominent in the mosses." "Antheridia of lycopods and horsetails each contain hundreds of angrogones, but these do not appear to be aggregated in cellular blocks."

339. (KC3.19) **Antheridium morphology: Naked (0); With Jacket (1).**
Rule 295. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
All embryophytes were scored as possessing a jacket (Kenrick and Crane 1997).

340. COMB (KC3.20, GRD5, RG1) **Antheridium development: Exogenous (0); Endogenous (1).**
Rule 296. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
The endogenous development of hornworts is thought to be unique (Kenrick and Crane 1997).

341. (S106) **Position of antheridia on gametophyte: Embedded (0); Partially or fully exposed (1).**
Rule 297. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
Little explicit information about this character for most taxa.

342. COMB (S110, GRD 6) **Antheridial stalk: Absent (0); Present (1).**
Rule 298. *Contingent on the presence of a multicellular antheridia (character 335, state 1).*
Following the scoring of Schneider *et al.* (2009), this character was deemed a feature of non-vascular plants, and so all vascular plants were scored absent. In Charales, the large basal cells of the antheridium is interpreted as a stalk (Garbary *et al.* 1993).

343. (S108) **Antheridial operculum: Absent** (**0**)**; Lateral, circular** (**1**)**; Terminal, circular** (**2**)**; Triangular** (**3**)**; Pore** (**4**)**.**
Rule 299*. Contingent on the presence of a multicellular antheridia (character 335, state 1).*

344. MOD (GRD7) **<u>Antheridial</u> operculum cells: Single** (**0**)**; Paired** (**1**)**; >2** (**2**)**; other** (**variable**) (**3**)**.**
Rule 300*. Contingent on the presence of a operculum (character 343, state 1).*

345. (GRD8, RG10) **Sperm in pollen tube: Absent** (**0**)**; Present** (**1**)**.**
Rule 301*. Contingent on the presence of a multicellular antheridia (character 335, state 1).*

346. (GRD9) **Jacket cells with chromoplasts: Absent** (**0**)**; Present** (**1**)**.**
Rule 302*. Contingent on the presence of a multicellular antheridia (character 335, state 1).*

347. (GRD10, RG11) **Number of sperm per male structure: 1** (**0**)**; 16 - 64** (**1**)**; 100 - 1000** (**2**)**; > 1000** (**3**)**; 2** (**4**)**; 4** (**5**)**.**
Rule 303*. Contingent on the presence of alternation of generations (character 3, states 2, 3).*

348. (KC3.23) **Paraphyses between gametangia: Absent** (**0**)**; Present** (**1**)**.**
Rule 304*. Contingent on the presence of alternation of generations (character 3, states 2, 3).*
Scored using M93 and S105. All higher and seed plants scored absent.

349 – 357. ***Rhizoids and endophytic associations***

349. NEW Rhizoids: Present (**0**)**; Absent** (**1**)**.**
Rule 305*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

350. (KC3.24) **Rhizoid cellularity: Unicellular** (**0**)**; Multicellular** (**1**)**.**
Rule 306*. Contingent on the presence of rhizoids (character 349, state 0).*
Ferns were all scored unicellular, based on their similarity to the unicellular root hairs of angiosperms (Takahashi 1961). Non-vascular plants scored according to M97.

352. (S102) **Obligate mycorrhizae in gametophytes: Absent** (**0**)**; Present** (**1**)**.**
Rule 307*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Absent in all seed plants and currently unknown in fossils. There is currently no evidence for mycorrhizae in mosses (Pressel *et al.* 2010), though there is in hornworts (Desiro *et al.* 2013) and liverworts (Ligrone *et al.* 2007). *Lycopodium* does possess a mycorrhizal gametophyte, whereas *Phylloglossum* undergoes an ontogenetic shift from mycorrhizal to photosynthetic (Whittier and Braggins 2000) and *Lycopodiella* is entirely photosynthetic (Whittier and Carter 2007).

353. NEW Fungal endophyte association with gametophyte: septate (**0**)**; aseptate** (**1**)**.**
Rule 308*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*

354. NEW Septate fungal endophyte association with gametophyte: Ascomycota (**0**)**; Basidiomycota** (**1**)**.**
Rule 309*. Contingent on the presence of septate fungal endophyte (character 353, state 0).*

355. NEW Aseptate fungal endophyte: Mucoromycotina (**0**)**; Glomeromycotina** (**1**)**; Mucoromycotina** + **Glomeromycotina** (**dual partnership**) (**2**)**.**
Rule 310*. Contingent on the presence of aseptate fungal endophyte (character 353, state 1).*

356. Mycoheterotrophy in gametophyte: Present (0); Absent (1).
Rule 311*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
Note that for some taxa e.g. *Lycopodiella* we will need to score multiple states.

357. Cyanobacterial endophytes: Absent (0); Present (1)
Rule 312*. Contingent on the presence of a multicellular gametophyte (character 300, state 1).*
V. Zoospores and Spermatozoids (358 - 453)

358 – 381. ***Zoospores, zoosporangia and flagellated vegetative cells.***

358. MOD (M15) **Zoospores: Absent (0); Present (1);** ~~**Present, flattened (2).**~~
All embryophytes were scored absent.

359. (M27) **Vegetative cell with flagella: Absent (0); Present (1).**
All embryophytes were scored absent.

360. (M24) **Number of flagella per vegetative cells or zoospores: Two (0); Four (1); One (2); More than four (3).**
Rule 313*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
All embryophytes scored not applicable. The original scoring was changed so that green algae lacking both zoospores and flagellated vegetative cells were scored not applicable, rather than a separate character state.

361. (M25) **Retraction of flagella during division: No (0); Yes (1).**
Rule 314*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable.

362. (M26) **Angle of basal bodies relative to motion: Angled (0); Perpendicular (1); Parallel (2).**
Rule 315*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable.

363. (M72) **Multilayered Structure (MLS): Absent (0); Present (1).**
Rule 316*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable. The MLS is a feature of all flagellated cells of charophytes (Renzaglia and Garbary 2001).

365. (M28) **Basal bodies distant via migration in development: No (0); Yes (1).**
Rule 317*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable.

366. (M29) **Flagella extend to right on motile cells: No (0); Yes (1).**
Rule 318*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*

This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable.

367. (M30) **Flagellar apparatus displaying 180 degree rotational symmetry: No (0); Yes (1).**
Rule 319. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable. The 180 degree symmetry and the X-2-X-2 flagellar apparatus was gained following the divergence of the charophycean and chlorophyceae algae (Melkonian 1982).

369. (M32) **Basal body overlap in motile cells: No (0); Yes (1).**
Rule 320. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable.

370. (M33) **Basal body core connection: Absent (0); Present (1).**
Rule 321. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable.

371. (M55) **Stigma: Absent (0); Present (1).**
Rule 322. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable. The stigma (or eyespot apparatus) is not present in charophytes (Goffinet and Shaw 2009), so all were scored absent.

372. (M57) **Apical insertion of flagella: Absent (0); Present (1).**
Rule 323. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* This character refers exclusively to zoospores or vegetative cells, rather than to motile sperm cells. All embryophytes were scored not applicable. All taxa possessing a MLS were scored absent, as the flagella insert into the MLS (Melkonian, 1981).

373. (M58) **Zoosporangia abscise: No (0); Yes (1).**
Rule 324. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* All taxa lacking zoospores scored not applicable.

374. (M59) **Zoosporangia operculate: No (0); Yes (1).**
Rule 325. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* All taxa lacking zoospores scored not applicable.

375. (M60) **Zoosporangial exit plug: No (0); Yes (1).**
Rule 326. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).* All taxa lacking zoospores scored not applicable.

376. (M61) **Keeled flagella: Absent (0); Present (1).**
Rule 327. *Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*

Following the scoring of Mishler *et al.* (1994), all charophytes were scored not applicable.

377. (M65) **Organic scales covering flagella: Absent** (**0**)**; Present** (**1**)**.**
Rule 328*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
Scales covering the flagella are a feature of some chlorophytes, but also a characteristic of charophytes (Garbary *et al.* 1993).

378. (M68) **SMAC or system I fiber: Absent** (**0**)**; Present** (**1**)**.**
Rule 329*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
Moestrup (1978) did not report any observations of striated fibers in charophytes, so all were scored absent.

379. (M70) **Distal fiber in motile cell: Absent** (**0**)**; Present** (**1**)**.**
Rule 330*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
All charophytes scored absent following the scoring of Mishler *et al.* (1994).

380. (M71) **Specialised zoosporangia: Absent** (**0**)**; Present** (**1**)**.**
Rule 331*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
All taxa lacking zoospores were scored not applicable.

381. (M14) **Vegetative cells or zoospores spindle shaped: No** (**0**)**; Yes** (**1**)**.**
Rule 332*. Contingent on the presence of zoospores /vegetative cells with flagella (character 358, state 1; character 359, state 1).*
All charophytes were scored not applicable.

382 – 385. ***Zygotes***

382. MOD (M49) **Enclosed zygote produced: No** (**Naked**) (**0**)**; Yes** (**1**)**.**
All extant embryophytes were scored (1) following the scoring of Mishler *et al.* (1994).

383. (M90) **Vertical division of zygote: Absent** (**0**)**; Present** (**1**)**.**
Scored in ferns using S126. A vertical first division of the zygote is a derived feature of hornworts and leptosporangiate ferns (Schneider *et al.* 2009; Ligrone *et al.* 2012). It also occurs in the Piperaceae and Loranthaceae (Gifford and Foster 1989). All other extant taxa were scored absent. Fossil taxa were scored unknown.

384. (M19) **Multiple sporulation/fission: Absent** (**0**)**; Present** (**1**)**.**
All embryophytes were scored absent.

385. (M16) **Autospores/colonies: Absent** (**0**)**; Present** (**1**)**.**
All embryophytes and charophytes were scored absent following the coding of Mishler *et al.* (1994).

386 – 393. ***Spermatogenous cells***

386. NEW Flagellated spermatogenous cells: Present (**0**)**; Absent** (**1**)**.**

Embryophytes lacking flagellated sperm cells include the angiosperms and most gymnosperms, excluding *Ginkgo* and cycads.

387. (GRD11) **Late spermatogenous cells shape: Angular** (**0**)**; Rounded** (**1**)**; Intermediate** (**2**)**.**
Rule 333*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

388. (GRD12, RGD5) **Nascent spermatids: paired** (**0**)**; Not paired** (**1**)**.**
Rule 334*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

389. (GRD14, RGD6) **Diagonal spindle in final mitosis division <u>of spermatogenous cells</u>: Present** (**0**)**; Absent** (**1**)**.**
Rule 335*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

390. (GRD15, RGD12) **Replication of the centrioles: Present** (**0**)**; Absent** (**1**)**.**
Rule 336*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

391. (GRD16, RGD13) **Time of origin of centrioles or centriole-generating organelles: Always present** (**0**)**; Spermatid mother cells** (**1**)**; Spermatid mother cell progenitor** (**2**)**; Earlier** (**in spermatogenous tissue**) (**3**)**.**
Rule 337*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

392. (GRD17, RGD14) **Basal bodies and flagella: Two** (**0**)**; More than two** (**1**)**.**
Rule 338*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

393. COMB (GRD18, RGD15) **Origin of centrioles: Right angles** (**0**)**; Bicentrioles** (**1**)**; Branched multiple** (**2**)**; Blepharoplast with templates** (**3**)**; Blepharoplast with centrioles** (**4**)**.**
Rule 339*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

394 – 430. ***Spermatozoid flagellar apparatus and cytoskeleton***

394. (GRD19, RGD28) **Basal body position: At right angles (at least during early development) (0); side-by-side (1); staggered/anterior/posterior (2); staggered/continous (3); staggered laterally (4).**
Rule 340*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

395. (GRD20) **Proximal extensions: Absent (0); Long (1); Short (2).**
Rule 341*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

396. (GRD22) **Stellate transition: Present (0); Absent (1).**
Rule 342*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

397. (GRD23, RGD18) **Transient cartwheel extensions: Absent (0); Present (1).**
Rule 343*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

398. (GRD24) **Connecting fibers between the basal bodies: Present (0); Absent (1); Fine filaments (2).**
Rule 344*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

399. (GRD25, RGD27) **Basal body structure: Monomorphic (0); Dimorphic (1).**
Rule 345*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

400. (GRD26) **Basal body staggering associated with growth of microtubules triplets: Absent (0); Present (1).**
Rule 346*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

401. (GRD27, RGD38) **Regression of longitudinal spline: Absent (0); Partial (1); Complete (2).**
Rule 347*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

402. (GRD28, RGD37) **Longitudinal spline / Anterior mitochondrion elongation in relation to the longitudinal axis of the spermatid: Parallel (0); Perpendicular (1).**
Rule 348*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*

Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

403. (GRD29, RGD40) **Change in the substructure of the longitudinal spline at maturity: Absent (0); Partially occluded (1); General loss of plate clarity (2); S_2 occluded (3); S_3 occluded (4); S_4 occluded (5).**
Rule 349*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

404. (GRD30, RGD43) **Spline aperture: Absent (0); Present (1).**
Rule 350*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

405. (GRD31, RGD43) **Spline aperture location: Left of centre (0); Right of centre (1).**
Rule 351*. Contingent on the presence of a spline aperture (character 404, state 1).*
Refers to spermatogenous cells only.

406. (GRD32, RGD16) **Position of developing MLS: Adjacent to basal bodies (0); beneath basal bodies (1).**
Rule 352*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

407. (GRD33) **Growth of longitudinal spline: No growth (0); Anterior (1); Posterior (2); Anterior and posterior (3).**
Rule 353*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

408. (GRD34, RGD17) **Stratified plaque between nascent blepharoplast: Absent (0); Present (1).**
Rule 354*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

409. (GRD35, RGD34) **Spline / longitudinal spline orientation: 90° (0); 45° (1); Variable (2); Absent (3); 28-45° (4); 16° (5).**
Rule 355*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

410. (GRD36) **Left-hand taper to spline: Absent (0); Present (1).**
Rule 356*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

411. (GRD37, RGD33) **Posterior notch to longitudinal spline: Absent** (**0**)**; Present** (**1**)**.**
Rule 357*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

412. (GRD38, RGD35) **Longitudinal spline position: Under all basal bodies** (**0**)**; Under anterior basal body only** (**1**)**; later some basal bodies** (**2**)**; absent** (**3**)**; to end of PBB** (**4**)**.**
Rule 358*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

413. (GRD39, RGD46) **Stray spline microtubules: Absent** (**0**)**; Present** (**1**)**; Develops late** (**2**)**.**
Rule 359*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

414. (GRD40), RGD47) **Accessory band of microtubules: Absent** (**0**)**; Present** (**1**)**.**
Rule 360*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

415. (GRD41, RGD41) **Increasing spline microtubules behind longitudinal spline: Absent** (**0**)**; Present** (**1**)**.**
Rule 361*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

416. (GRD42) **Longitudinal spline wider than spline: Absent to slight** (**0**)**; Extensive** (**1**)**.**
Rule 362*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

417. (GRD43) **Spline growth: Posterior** (**0**)**; Anterior/Posterior** (**1**)**.**
Rule 363*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

418. (GRD44) **Maturational elongation of anterior mitochondrion: Absent** (**0**)**; Posterior** (**1**)**.**
Rule 364*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

419. (GRD45 RGD42) **Spline widths** (**in the mature cell**)**: 40** (**0**)**; 50 - 110** (**1**)**; 150 - 180** (**2**)**; approx. 200** (**3**)**; approx. 300** (**4**)**;** >**1000** (**5**)**.**
Rule 365*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*

Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

420. (GRD46) **Spline shank: tapering uniformly** (**0**)**; tapering right side** (**1**)**.**
Rule 366*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

421. (GRD47) **Spline shank width: wide** (**0**)**; less than 4 tubules** (**1**)**.**
Rule 367*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

422. (GRD48) **Fibrous sheath at maturity: Absent** (**0**)**; Present** (**1**)**.**
Rule 368*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

423. (GRD49) **Osmiophilic crest: Absent** (**0**)**; Present** (**1**)**.**
Rule 369*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

424. (GRD50) **Anterior osmiophilic ridge: Absent** (**0**)**; Present** (**1**)**.**
Rule 370*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

425. (GRD51, RGD30) **Changes in basal bodies at maturity: Absent** (**0**)**; Dense material at extreme tip** (**1**)**; Basal body cartwheel with plug of matrix** (**2**)**; Basal body triplets impregnated with matrix** (**3**)**.**
Rule 371*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

426. COMB (GRD52, RG29) **Matrix around basal bodies: Homogeneous** (**0**)**; Mottled** (**1**)**; Amorphous** (**2**)**.**
Rule 372*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

427. (RGD19) **Consistency of structure between basal bodies: Undifferentiated** (**0**)**; Absent** (**1**)**; Fine filaments with centrin** (**2**)**.**
Rule 373*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

428. MOD (GRD53) **Position of the stellate pattern: Partially extracentriolar (0); Entirely intracentriolar (1); Both (2).**
Rule 374*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

429. MOD (GRD55) **Late blepharoplast with transient core: Yes (0); No (1).**
Rule 375*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

430. (GRD56) **Flagella number: 2 (0); 40 - 150 (1); >1000 (2).**
Rule 376*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

431 – 442. ***Spermatozoid nucleus***

431. (GRD58) **Nuclear shape at maturity: ovoid (0); elongate (1).**
Rule 377*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

432. (GRD59) **Nuclear posterior shape: not expanded (0); expanded (1).**
Rule 378*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

433. (GRD60) **Condensed chromatin: Homogeneous (0); Heterogeneous (1); Absent (2).**
Rule 379*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

434. (GRD61) **Median constriction: Absent (0); Present (1).**
Rule 380*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

435. (GRD62) **Overlap of anterior mitochondrion and nucleus: Absent (0); Present (1).**
Rule 381*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

436. (GRD63) **Spline attached to nucleus: Attached at maturity (0); Detached at maturity (1); Never attached (2).**
Rule 382*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*

Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

437. (GRD64) **Spline growth and nuclear shaping: Absent (0); Present (1).**
Rule 383*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

438. (GRD65) **Direction of nuclear compaction: outer shell (0); anterior to posterior (1); at equal rates along nucleus (2); general increase in density (3).**
Rule 384*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

439. (GRD66) **Condensed chromatin strands: Spaghetti-like (0); Perpendicular to spline (1); Spiral/central strand (2); General compaction (3); Spikes (4); Irregular plates (5); Solid mass from anterior tip (6).**
Rule 385*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

440. (GRD67) **Diverticulum during shaping: Absent (0); Present (1).**
Rule 386*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

441. (GRD68) **Excess nuclear envelope loss: Absent (0); After condensation (1); Gradual posteriorly (2); Gradual throughout (3).**
Rule 387*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

442. (GRD69) **Number of gyres of nucleus: Not coiled (0); 0.5 - 3 (1); >3 (2).**
Rule 388*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

443 – 450. ***Spermatozoid mitochondia***

443. MOD (GRD70) **Dense body in anterior mitochondrion: Absent (0); Present (1).**
Rule 389*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

444. (GRD72) **Mitochondria with plastids in young spermatids: Absent (0); Present (1).**
Rule 390*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*

Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

445. (GRD73) **Mitochondria with plastids in mature sperm: Absent** (**0**)**; Present** (**1**)**.**
Rule 391*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

446. (GRD74) **Specialized anterior mitochondrion: Present** (**0**)**; Absent** (**1**)**.**
Rule 392*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

447. (GRD75) **Additional mitochondria in anterior of cell: Absent** (**0**)**; Row of mitochondria behind anterior mitochondrion** (**1**)**; numerous unspecialized** (**2**)**.**
Rule 393*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

448. (GRD77) **Osmiophilic material with anterior mitochondrion: Absent** (**0**)**; Present** (**1**)**.**
Rule 394*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

449. (GRD78) **Number of mitochondria in sperm: Many** (**0**)**; Two** (**1**)**.**
Rule 395*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

450. (GRD79) **Cristae sacs to baffles: Absent** (**0**)**; Present** (**1**)**.**
Rule 396*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

451 – 455. ***Spermatozoid plastids***

451. (GRD82) **Plastid/nuclear association: Absent** (**0**)**; Present** (**1**)**.**
Rule 397*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

452. (GRD83) **Starch grains in single plastid: More than one** (**0**)**; one** (**1**)**.**
Rule 398*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

453. (GRD84) **Sperm plastid contacting nucleus: Absent** (**0**)**; Present** (**1**)**; Present via chloroplast extension** (**2**)**.**
Rule 399*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

454. (GRD85) **Location of plastids: Posterior** (**0**)**; Central** (**1**)**; Posterior but not attached to spline** (**2**)**.**
Rule 400*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

455. (GRD86) **Fibrillenscheide: Absent** (**0**)**; Present** (**1**)**.**
Rule 401*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable. This feature is unique to liverworts.

456 – 457. ***Sperm maturation***

456. (GRD89) **Cytoplasmic loss: Absent** (**0**)**; Partial** (**1**)**; Complete or with tiny remnant** (**2**)**.**
Rule 402*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

457. (GRD90) **Sperm symmetry: Longitudinal symmetry** (**0**)**; Asymmetrical** (**1**)**.**
Rule 403*. Contingent on the presence of flagellated spermatogenous cells (character 386, state 0).*
Embryophytes lacking flagellated sperm cells (angiosperms and most gymnosperms, excluding *Ginkgo* and cycads) were scored not applicable.

VI. Spores, pollen, embryology and seeds

458 – 481. ***Spores***

458. NEW Megaspore /Spore tetrads: Present (**0**)**; Absent** (**1**)**.**
Rule 404*. Contingent on alternation of generations (character 3, states 2, 3).*

459. NEW Permanently fused tetrads: Present (**0**)**; Absent** (**1**)**.**
Rule 405*. Contingent on alternation of generations (character 3, states 2, 3).*

460. MOD (HB76) **Tetrad arrangement: Multiplanar** (**0**)**; Uniplanar** (**1**)**; Monolete** (**2**)**.**
Rule 406*. Contingent on the presence of spore tetrads (character 458, state 0).*
Refers to both megaspore and homospore tetrads. All homosporous taxa were scored according to (Schneider *et al.* 2009) or unknown. Following the scoring of HB and S, all angiosperms were scored linear.

461. NEW Spore dyads: Present (**0**)**; Absent** (**1**)**.**
Rule 407*. Contingent on alternation of generations (character 3, states 2, 3).*

462. NEW Permanently fused dyads: Present - fused (0); Absent - naturally separating (1).
Rule 408*. Contingent on the presence of spore dyads (character 461, state 0).*

463. NEW Multicellular spores: Present (1); Absent (0).
Rule 409*. Contingent on alternation of generations (character 3, states 2, 3).*

464. (HB88) **Nuclei per microgametophyte: Five or more (0); Four, with one tube nucleus formed via the second division (1); Four with one tube nucleus formed via the first division (2); Three, generative division pre-pollination (3); Two, generative division post-pollination (4).**
Rule 410*. Contingent on heterosporous sporogenesis (character 4, state 1).*
All taxa were scored using S120 and DE80.

465. (M94) **Perine on spores: Absent (0); Solid perine present (1); Reduced perine, orbicules and Ubisch bodies present (2).**
Rule 411*. Contingent on alternation of generations (character 3, states 2, 3).*
Scored using KC3.13 and S87. Though seed plants are potentially lacking a perine, Blackmore (1990) argue that this structure is homologous to the orbicules found in seed plants. All seed plants were scored as possessing orbicules. Following (Kenrick and Crane 1997), all fossil taxa were scored as unknown, as the perine does not preserve well. *Phaeoceros* was scored unknown following the possible presence of a perine (Villarreal and Renzaglia 2006).

466. (S88) **Perispore/perine prominence relative to exine/exospore: Not prominent (0); Prominent (1).**
Rule 412*. Contingent on the presence of a perispore (character 465, states 1, 2).*
The wording was changed from the original character as absent is already a character state in S86. This character was applied only to character that possess a solid perine. The perispore of lycophyte spores were scored as not prominent, based on their description as 'thin and compact' (Tryon and Lugardon 1991).

467. (S89) **Perine /Perispore surface: Smooth or plain (0); Obviously patterned (1).**
Rule 413*. Contingent on the presence of a perispore (character 465, states 1, 2).*
This character was applied only to character that possess a solid perine. *Lycopodium* was scored following Pryer *et al.* (1995).

468. (S86) **Spores with acrolamella: Absent (0); Present (1).**
Rule 414*. Contingent on alternation of generations (character 3, states 2, 3).*
This is a feature of Marsileaceae and Salviniaceae (Schneider *et al.* 2009).

469. (S91) **Exospore surface: Smooth or plain (0); Sculptured (1).**
Rule 415*. Contingent on alternation of generations (character 3, states 2, 3).*
This character was altered from the original to include only exospores, as the ornamentation of pollen surfaces was captured in HB85. Consequently, all seed plants were scored not applicable. Lower vascular plants were scored according to descriptions of (Kenrick and Crane 1997).

470. (HB83) **Exine striations: Absent (0); Present (1).**
Rule 416*. Contingent on heterosporous sporogenesis (character 4, state 1).*

This character was interpreted to refer only to heterosporous taxa, and so all homosporous taxa were scored not applicable. The scoring of angiosperms as performed using character DE90, under the assumption that striate muri are equivalent to exine striations.

471. COMB (HB84, DE88) **Tectum: Continuous to microperforate** (**0**)**; Perforate to semitectate** (**1**)**; Reduced** (**2**)**.**
Rule 417*. Contingent on heterosporous sporogenesis (character 4, state 1).*
Oryza was scored as continuous due to its smooth surface with no reticulate cavities (Chaturvedi *et al.* 1998). *Glycine* was scored as unknown.

472. (HB85) **Supratectural spinules: Absent** (**0**)**; Present** (**1**)**.**
Rule 418*. Contingent on heterosporous sporogenesis (character 4, state 1).*
These are also described as microechinate ornamentation. All angiosperms were scored following DE91. No reference was made to spinules on either *Glycine* or *Oryza*, but both were scored unknown.

473. (HB87) **Endexine: Uniformly thick, laminated** (**0**)**; Absent** (**1**)**; Thin, except under apertures** (**2**)**.**
Rule 419*. Contingent on heterosporous sporogenesis (character 4, state 1).*
All angiosperms were scored according to DE94, with the assumption that a laminate endexine is found only in gymnosperms (Doyle 2006).

474. (S85) **Spore laesura: Linear** (**0**)**; Triradiate** (**1**)**; Circular** (**2**)**; Sulcate** (**3**)**.**
Rule 420*. Contingent on the presence of spore tetrads (character 458, state 0) or dyads (character 461, state 0).*
All seed plants were excluded from this character as their diversity was captured in DE85. Lower vascular plants were scored according to descriptions of Kenrick and Crane (1997), following the notion that a trilete scar = triradiate.

475. (S92) **Pseudoendospore: Absent** (**0**)**; Present** (**1**)**.**
Rule 421*. Contingent on alternation of generations (character 3, states 2, 3).*
This feature is reported as a feature of ferns, that is not homologous to the intine/endospore of other taxa (Tryon and Lugardon 1991; Schneider *et al.* 2009).

476. (S93) **Paraexospore: Absent** (**0**)**; Present** (**1**)**.**
Rule 422*. Contingent on alternation of generations (character 3, states 2, 3).*
This feature is reported as unique to *Isoetes* and *Selaginella* (Schneider *et al.* 2009).

477. (S94) **Spore wall development: Centripetal** (**0**)**; Centrifugal** (**1**)**.**
Rule 423*. Contingent on alternation of generations (character 3, states 2, 3).*
All fossil taxa were scored unknown. All other taxa were scored according to Schneider *et al.* (2009) and Wallace *et al.* (2011). The unusual development of Andreaea has recently been highlighted and so it was scored unknown.

478. (DE78) **Tapetum type: Secretory** (**0**)**; Amoeboid** (**1**)**.**
Rule 424*. Contingent on alternation of generations (character 3, states 2, 3).*
All ferns were scored according to S95. All fossil taxa were scored unknown. The presence or absence of Ubisch bodies in seed plants was used a proxy, since they are only associated with secretory tapeta. Tapeta are absent in some gymnosperms, and so they were scored not applicable (Biswas and Johnri 1997). The secretory type is found through non-vascular plants.

480. (S97) **Chlorophyllous spores: Absent** (**0**)**; Present** (**1**)**.**
Rule 425*. Contingent on alternation of generations (character 3, states 2, 3).*
These spores occur in a few families of ferns (Lloyd and Klekowski 1970).

481. (S85) **Spore germination pattern: Equatorial** (**0**)**; Polar** (**1**)**; Amorphous** (**2**)**.**
Rule 426*. Contingent on alternation of generations (character 3, states 2, 3).*
Difficult to score as there is not much information – there were a lot of unknowns in the original scoring and it may be that all currently unscored taxa will be left unknown.

482 – 495. ***Pollen***

482. (DE84) **Microspore aperture type: Polar, including sulcate, ulcerate and disulcate** (**0**)**; Inaperturate** (**1**)**; Sulculate** (**2**)**;** (**syn**)**Tricolpate, with colpi arranged according to Garside's law, with or without alternating colpi** (**3**)**; Tricolpate** (**4**)**; Proximal tetrad scar** (**5**)**.**
Rule 427*. Contingent on heterosporous sporogenesis (character 4, state 1).*
All seed plants scored according to HB79. Non-seed plant microspores were scored based on descriptions of Kenrick and Crane (1997).

483. (DE85) **Distal aperture shape: Elongate** (**0**)**; Round** (**1**)**.**
Rule 428*. Contingent on the presence of a polar aperture (character 482, state 0).*
Taxa with multiple or no apertures were scored not applicable. Most seed plants possessed elongate apertures. Among extant gymnosperms, the conifers and *Welwitschia* were elongate (Millay and Taylor 1974; Webb and Moore 1978; Stockey 1980; Rydin and Friis 2005), whereas the large apertures of *Cycas* and *Ginkgo* appear rounded (Fernando *et al.* 2010). Extinct seed ferns were largely scored according to descriptions of Taylor *et al.* (2009), and all possessed elongate apertures. The Bennettitales also possessed elongate apertures (Zavialova *et al.* 2009). Taxa lacking a clear description (e.g. *Pentoxylon* = "broad") were scored unknown.

484. (DE86) **Distal aperture branching: Unbranched** (**0**)**; With several branches** (**1**)**.**
Rule 429*. Contingent on the presence of a polar aperture (character 482, state 0).*
Following the references of character 482 (DE85), none of the seed plant taxa were described as possessing a branched aperture, and so all non-angiosperm taxa scored for 479 were scored as unbranched.

485. (HB86) **Aperture membrane: Smooth or weakly sculptured** (**0**)**; Sculptured** (**1**)**.**
Rule 430*. Contingent on heterosporous sporogenesis (character 4, state 1).*
All angiosperms were scored following DE93. Both *Oryza* and *Glycine* were scored unknown.

486. (HB81) **Saccate pollen: Absent** (**0**)**; Present** (**1**)**.**
Rule 431*. Contingent on heterosporous sporogenesis (character 4, state 1).*
Saccate pollen is a feature only of some wind pollinated conifers, and so all other taxa were scored absent.

487. (DE81) **Pollen unit: Monads** (**0**)**; Tetrads** (**1**)**.**
Rule 432*. Contingent on heterosporous sporogenesis (character 4, state 1).*
Though appear to be the predominant form, taxa lacking an explicit description of the pollen unit according to Taylor *et al.* (2009) were scored unknown.

488. (DE82) **Pollen size, average: Large, > 50 µm (0); Medium, 20 – 50 µm (1); Small, < 20 µm (2).**
Rule 433*. Contingent on heterosporous sporogenesis (character 4, state 1).*
Average values were not always available and so the approximate sizes supplied by Taylor *et al.* (2009) were used. For extant gymnosperms, the palynological database of the University of Arizona was used, as well as measurements obtained by Knight *et al.* (2010). *Glycine* was scored according to Koti *et al.* (2004).

489. (DE83) **Pollen shape: Boat-shaped to elliptical (0); Globose (1); Triangular to angulaperturate (2); Bisaccate (3); Monosaccate (4).**
Rule 434*. Contingent on heterosporous sporogenesis (character 4, state 1).*
This character was expanded from the original to incorporate gymnosperms. Descriptions of Taylor *et al.* (2009) were used for all extinct taxa, along with HB88. *Welwitschia* and *Ephedra* were determined as elliptical (Rydin and Friis 2005), while *Gnetum* is spherical (Gillespie and Nowicke 1994).

490. COMB (DE88, HB82) **Infratectum structure: Massive/spongy alveolar (0); Honeycomb alveolar (1); Granular (2); Intermediate (3); Columellar (4).**
Rule 435*. Contingent on heterosporous sporogenesis (character 4, state 1).*

491. (DE89) **Grading of reticulum: Uniform (0); Finer at the end of sulcus (1); Finer at poles (2).**
Rule 436*. Contingent on a honeycomb alveolar infratectum (character 490, state 1).*
This taxa was only scored for taxa possessing a honeycomb alveolar infratectum.

492. (DE92) **Prominent spines: Absent (0); Present (1).**
Rule 437*. Contingent on heterosporous sporogenesis (character 4, state 1).*
Non-angiosperm seed plants were scored absent, due to no mention of prominent spines in any descriptions of Taylor *et al.* (2009).

493. (DE94) **Extra-apertural nexine stratification: Foot layer, not consistently foliated, no distinctly staining endexine or only problematic.traces (0); Foot layer and distinctly staining endexine, or endexine only (1); all or in part foliated, not distinctly staining (2).**
Rule 438*. Contingent on heterosporous sporogenesis (character 4, state 1).*
All or in part foliated, not distinctly staining. This character was interpreted to apply only to taxa which feature thin nexine except around apertures (see HB87). Consequently, all taxa with uniformly thick endexine were scored not applicable.

494. (DE95) **Nexine thickness: Absent or discontinuous (0); Thin but continuous (1); Thick, 1/3 or more of exine (2).**
Rule 439*. Contingent on heterosporous sporogenesis (character 4, state 1).*
All taxa that were scored as having a thick, laminated endexine (HB87) were scored here as having a thick nexine.

495. (DE77) **Connective hypodermis: Unspecialised (0); Sclerenchymatous or endothecial (1).**
Rule 440*. Contingent on heterosporous sporogenesis (character 4, state 1).*
Very few references to this structure are made beyond angiosperms. All non-seed plants scored not applicable, remaining seed plants scored unknown.

496 – 552. ***Embryology and Seeds***

496. (M20) **Type of sex: Isogamy** (**0**)**; Anisogamy** (**1**)**; Oogamy** (**2**)**.**
All embryophytes as well as *Chara* and *Coleochaete* were scored as oogamous (Doyle 2013).

497. (M82) **Embryo: No** (**0**)**; Yes** (**1**)**.**
All extant embryophytes and those with preserved gametophytes were scored (1).

498. (S125) **Product of fertilization: Diploid zygote and embryo** (**0**)**; Diploid zygote and embryo, plus triploid endosperm** (**1**)**.**
Rule 441*. Contingent on oogamy (character 496, state 2).*
All angiosperms were scored as having a triploid endosperm. Even in taxa that predominantly feature a perisperm, the endosperm is initiated (Friedman *et al.* 2008). Taxa that are not oogamous were scored not applicable.

499. (HB100) **Feeder in embryo: Absent** (**0**)**; Present** (**1**)**.**
Rule 442*. Contingent on oogamy (character 496, state 2).*
The feeder is a unique structure of *Welwitschia* and *Gnetum* (Crane 1985), and so all other taxa were scored absent.

500. (S127) **Embryo orientation: Exoscopic** (**0**)**; Endoscopic** (**1**)**; Prone** (**2**)**.**
Rule 443*. Contingent on oogamy (character 496, state 2).*
As the embryo develops within the seed, all seed plants were scored endoscopic (Bell and Hemsley 2000). By contrast, orientation in mosses, liverworts and hornworts is exoscopic (Bell and Hemsley 2000).

501. (S128) **Suspensor: Absent** (**0**)**; Present** (**1**)**.**
Rule 444*. Contingent on oogamy (character 496, state 2).*
Schneider *et al.* (2009) describe the suspensor as a fixed condition in certain land plants, and so lineages where they are reported, such as gymnosperms and angiosperms, were all scored present.

502. (S130) **Embryo development: Derived from a single, uninucleate cell** (**0**)**; Derived from several free nuclei** (**1**)**.**
Rule 445*. Contingent on oogamy (character 496, state 2).*
This character emphasises the unique free-nuclear phase present in gymnosperms (Gifford and Foster 1989), and so all other taxa were scored as derived from a uninucleate cell.

503. (M74) **Oogonium associated with sterile cells: No** (**0**)**; Yes** (**1**)**.**
All extant embryophytes and those with preserved gametophytes were scored (1).

504. (M75) **Eggs retained in oogonium: No** (**0**)**; Yes** (**1**)**.**
All extant embryophytes and those with preserved gametophytes were scored (1).

505. (M78) **Zygote retained: No** (**0**)**; Yes** (**1**)**.**
All extant embryophytes and those with preserved gametophytes were scored (1).

506. NEW Placenta: Absent (**0**)**; Present** (**1**)**.**

507. (M79) **Placental transfer cells: No** (**0**)**; Yes** (**1**)**.**
Most extant embryophytes and those with preserved gametophytes were scored (1).

508. NEW Position of placental transfer cells: sporophyte only (**0**)**; gametophyte only** (**1**)**; both generations** (**2**)**; sporophytic haustoria** (**3**)**.**
Rule 446*. Contingent on the presence of placental transfer cells (character 507, state 1).*

509. (HB98) **Proembryo: Not tiered** (**0**)**; Tiered** (**1**)**.**
Rule 447*. Contingent on oogamy (character 496, state 2).*
Doyle (1996) lists this character as a synapomorphy of derived gymnosperms, and so all other taxa were scored absent.

510. (HB99) **Secondary suspensor: Present** (**0**)**; Absent** (**1**)**.**
Rule 448*. Contingent on the presence of a suspensor (character 501, state 1).*
All taxa lacking a suspensor were scored not applicable. This character is absent in angiosperms (Doyle 1996).

511. (**HB94**) **Double fertilisation: Fusion of only one sperm with female gametophyte nucleus** (**0**)**; Regular fusion of both sperm** (**1**)**.**
All fossil taxa are scored as missing. The derived character state (1) of both of the Doyle (1996) characters 83 and 84 which are also HB characters 94 and 95 'constitute classic double fertilization of angiosperm type' (Doyle 1996). Thus all non-angiosperm or gnetales taxa are scored as (0) (see details of character 83 and 84 in Doyle (1996).

512. (M6) **Pollen tube. Absent** (**0**)**; Present** (**1**)**.**

513. (HB60) **Seed: Absent** (**0**)**; Present** (**1**)**.**
Scored according to M108 and S82. All non-seed plants scored absent.

514. (HB61) **Anatomical symmetry of ovule: Radial** (**0**)**; 180 degree rotational** (**1**)**; Bilateral** (**2**)**.**
Rule 449*. Contingent on the presence of a seed (character 513, state 1).*
Following Hilton and Bateman (2006) and Endress (2011), the symmetry in angiosperms and gymnosperms is not comparable and all angiosperms were scored unknown.

515. (DE113) **Placentation: Ventral** (**0**)**; Laminar-diffuse or dorsal** (**1**)**.**
Rule 450*. Contingent on the presence of a seed (character 513, state 1).*
As the ovules of gymnosperms are naked (Biswas and Johnri 1997), this character was scored not applicable for all non-angiosperms. *Glycine* and *Oryza* were scored as unknown, as their marginal and basal placentation were not easily applied with the current character states.

516. (DE114) **Ovule direction: Pendent** (**0**)**; Horizontal** (**1**)**; Ascendent** (**2**)**.**
Rule 451*. Contingent on the presence of a seed (character 513, state 1).*
This character was scored not applicable for all non-angiosperms. General descriptions of the Fabaceae state a pendulous to ascendant ovule (Simpson 2011), and so *Glycine* was left unknown.

517. (DE121) **Chalaza: Unextended** (**0**)**; Pachychalazal** (**1**)**; Perichalazal** (**2**)**.**
Rule 452*. Contingent on the presence of a seed (character 513, state 1).*

The nature of the chalaza in this context was unknown for non-angiosperms (Wang 2010), with the exception of extant conifers, which were scored pachychalazal by (Nixon *et al.* 1994).

518. (HB63) **Integuments: Lobate preintegument** (**0**)**; With simple apex** (**1**)**; With bifid apex** (**2**)**; With straight, tubular micropyle** (**3**)**; With inverted micropyle relative to strobilus axis** (**4**)**.**
Rule 453*. Contingent on the presence of a seed (character 513, state 1).*
Following the scoring of Hilton and Bateman (2006), and given that there is no reference to a bifid apex in angiosperms, all angiosperms were scored as having a simple apex.

519. (DE116) **Number of integuments: Two** (**0**)**; One** (**1**)**; Three** (**2**)**.**
Rule 454*. Contingent on the presence of a seed (character 513, state 1).*
This character was expanded to reflect the bracteolate second and third later in *Gnetum* and *Welwitschia* (Stevenson 2012).

520. (HB62) **Fusion of integuments: Free** (**0**)**; Fused for more than half of its length** (**1**)**.**
Rule 455*. Contingent on the presence of a seed (character 513, state 1).*
Following the scoring of Hilton and Bateman (2006) and (Doyle 2006), all angiosperms were coded free.

521. (DE117) **Outer integument shape: Semiannular** (**0**)**; Annular** (**1**)**.**
Rule 456*. Contingent on the presence of a seed (character 513, state 1).*
Orthotropous taxa were scored unknown. *Glycine* and *Oryza* were scored unknown. All non-angiosperms were scored not applicable. As the outer integument of *Gnetum* and *Welwitschia* are fundamentally different, they were scored unknown.

522. (DE118) **Outer integument lobation: Unlobed** (**0**)**; Lobed** (**1**)**.**
Rule 457*. Contingent on the presence of a seed (character 513, state 1).*
Glycine and *Oryza* were scored unknown. All non-angiosperms were scored not applicable. As the outer integument of *Gnetum* and *Welwitschia* are fundamentally different, they were scored unknown.

523. (DE119) **Outer integument thickness, at the middle of the integument length: Two cells** (**0**)**; Two, three and four** (**1**)**; Four, five and more** (**2**)**.**
Rule 458*. Contingent on the presence of a seed (character 513, state 1).*
Glycine and *Oryza* were scored unknown. All non-angiosperms were scored not applicable. As the outer integument of *Gnetum* and *Welwitschia* are fundamentally different, they were scored unknown.

524. (DE120) **Inner integument thickness: Two cells** (**0**)**; Two, or two and three** (**1**)**; Three or more** (**2**)**.**
Rule 459*. Contingent on the presence of a seed (character 513, state 1).*
As the inner integument of angiosperms and Gnetales is thought to be homologous to the integument of other seed plants, where possible all seed plants were scored for this character. Extant gymnosperms were scored following Wang (2010) and fossil seed plants were scored according to the taxon descriptions of Taylor *et al.* (2009).

525. (HB65) **Integumentary apex sealing post-pollination: Absent** (**0**)**; Present** (**1**)**.**
Rule 460*. Contingent on the presence of a seed (character 513, state 1).*
Following the scoring of Hilton and Bateman (2006) and Doyle (2006), all angiosperms were scored present.

526. (HB70) **Integumentary vascularization: Numerous bundles, one in each lobe** (**0**)**; Two bundles dividing in a major plane** (**1**)**; Unvascularised** (**2**)**.**
Rule 461*. Contingent on the presence of a seed (character 513, state 1).*
Following the scoring of Hilton & Bateman (2006), all angiosperms were scored unknown. Endress (2011) reports that in some taxa, especially those with large ovules, there is vascularization of the integument, but the nature of this vascularization is unknown.

527. (HB66) **Salpinx: Absent** (**0**)**; Present** (**1**)**.**
Rule 462*. Contingent on the presence of a seed (character 513, state 1).*
Following the scoring of Hilton and Bateman (2006) and Doyle (2006), all angiosperms were scored present.

528. (DE122) **Nucellus: Crassinucellar, including weakly so** (**0**)**; Tenuinucellar** (**1**)**.**
Rule 463*. Contingent on the presence of a seed (character 513, state 1).*
Though the nucellus of extant gymnosperms is defined as crassinucellar Singh (2006) this character is not described in fossil seed plants, and so they were scored unknown.

529. (HB67) **Pollen chamber: Absent** (**0**)**; With a membranous floor** (**1**)**; Without a membranous floor** (**2**)**.**
Rule 464*. Contingent on the presence of a seed (character 513, state 1).*
Following Hilton and Bateman (2006), all angiosperms were scored absent.

530. (HB68) **Central column: Present** (**0**)**; Absent** (**1**)**.**
Rule 465*. Contingent on the presence of a seed (character 513, state 1).*
Following Hilton and Bateman (2006) all angiosperms were scored absent.

531. (HB69) **Pollen chamber sealing post pollination: Not sealed** (**0**)**; Sealed** (**1**)**.**
Rule 466*. Contingent on the presence of a pollen chamber (character 529, states 1, 2).*
All taxa scored as lacking a pollen chamber were scored not applicable.

532. (HB73) **Testa: Multiplicative** (**0**)**; Non-multiplicative** (**1**)**.**
Rule 467*. Contingent on the presence of a seed (character 513, state 1).*
All angiosperms were scored according to DE127.

533. (DE128) **Exotesta: Unspecialized** (**0**)**; Palisade or short sclerotic cells** (**1**)**; Tabular** (**2**)**; Longitudinally elongated, lignified cells** (**3**)**.**
Rule 468*. Contingent on the presence of a seed (character 513, state 1).*
All non-angiosperms were scored according to HB74.

534. (DE129) **Mesotesta lignification: Unlignified** (**0**)**; With a sclerotic layer** (**1**)**; With a fibrous layer** (**2**)**.**
Rule 469*. Contingent on the presence of a seed (character 513, state 1).*
The term mesotesta is rarely applied to gymnosperms, though here we treat it as equivalent to the middle 'sclerotesta' (Kozlowski 1972), an intermediate sclerified layer. Consequently, all taxa with a sclerotesta were scored as having a sclerotic mesotesta. Extinct seed plants were scored following descriptions of Taylor *et al.* (2009).

535. (HB64) **Sarcotesta: Absent** (**0**)**; Present** (**1**)**.**
Rule 470*. Contingent on the presence of a seed (character 513, state 1).*

While Hilton and Bateman (2006) scored angiosperms as absent, some angiosperms are reported as possessing a mesotesta that has formed into a fleshy sarcotesta (Doyle and Endress 2010).

536. (DE131) **Endotesta: Unspecialized (0); Single layer with fibrous reticulum (1); Multiple layers with fibrous reticulum (2); Tracheidal (3); Palisade of thick walled cells (4).**
Rule 471*. Contingent on the presence of a seed (character 513, state 1).*
Gymnosperms are all reported as having an endotesta that is thin, multi-layered and paper-like (Biswas and Johnri 1997). It is possible that this corresponds to the fibrous reticulum, but it was uncertain, so all non-angiosperms were scored unknown.

537. (DE133) **Ruminations: Absent (0); Testal (1); Tegminal and/or chalazal (2).**
Rule 472*. Contingent on the presence of a seed (character 513, state 1).*
All non-angiosperms scored according to HB75. As none were reported for *Oryza* (Bechtel and Pomeranz 1977) or *Glycine* (Miller *et al.* 1999), each was scored absent.

538. (DE123) **Fruit wall: Wholly or partly fleshy (0); Dry (1).**
Rule 473*. Contingent on the presence of flowers (character 227, state 1).*
This character applies only to angiosperms. Following general descriptions, both *Oryza* and *Glycine* were scored dry (Simpson 2011).

539. (DE124) **Lignified endocarp: Absent (0); Present (1).**
Rule 474*. Contingent on the presence of flowers (character 227, state 1).*
All taxa scored as dry in DE123 were scored unknown.

540. (DE125) **Fruit dehiscence: Indehiscent or dehiscing irregularly, dorsally or laterally (0); Dehiscence ventral or both ventral and dorsal (1); Horizontal dehiscence with vertical extensions (2).**
Rule 475*. Contingent on the presence of flowers (character 227, state 1).*
Oryza and other grasses possess a caryopsis, a dry indehiscent fruit (Dahlgren *et al.* 1985). *Glycine* has a pod which dehisces along dorsal and ventral sutures (Christiansen *et al.* 2002).

541. (DE126) **Hooked hairs on fruits: Absent (0); Present (1).**
Rule 476*. Contingent on the presence of flowers (character 227, state 1).*
Both *Oryza* and *Glycine* were scored absent.

542. (DE132) **Tegmen: Unspecialised (0); Both endo- and ectotegmen thick-walled (1); Exotegmen thick-walled to sclerotic (2).**
Rule 477*. Contingent on the presence of a seed (character 513, state 1).*
Taxa that were scored unitegmic (one integument) for character DE116 were scored not applicable as unitegmic seeds strictly have a testa (Schmid 1986). Gnetales were scored unknown.

543. (DE134) **Seed operculum: Absent (0); Present (1).**
Rule 478*. Contingent on the presence of a seed (character 513, state 1).*
This character appears to be a synapomorphy of the Nymphaeales (Rudall *et al.* 2008) and so *Oryza* and *Glycine* were scored absent. As this character is not described in any non-angiosperm ovule (Biswas and Johnri 1997; Taylor *et al.* 2009), all were scored absent.

544. (DE135) **Aril: Absent** (**0**)**; Present** (**1**)**.**
Rule 479*. Contingent on the presence of a seed (character 513, state 1).*
Beyond the scoring of Doyle and Endress (2010), there is no reference to an aril in any taxa besides Taxaceae (Biswas and Johnri 1997), and so all other taxa were scored absent.

545. (HB96) **Nutritive tissue in seed: Gametophytic** (**0**)**; Endosperm/perisperm** (**1**)**.**
Rule 480*. Contingent on the presence of a seed (character 513, state 1).*
Nourishment of the developing sporophyte via an endosperm/perisperm is a major innovation of angiosperms (Doyle 2013), and so all angiosperms were scored (1).

546. (DE137) **Endosperm development: Cellular** (**0**)**; Nuclear** (**1**)**; Helobial** (**2**)**.**
Rule 481*. Contingent on the presence of an endosperm (character 545, state 1).*
Only taxa scored as having endosperm in HB96 were scored for this character. The endosperm of *Glycine* forms via a free-nuclear stage that is subsequently cellularized (Chamberlin *et al.* 1994), as does *Oryza* (Brown *et al.* 1996).

547. (DE138) **Endosperm in the mature seed: Present** (**0**)**; Absent** (**1**)**.**
Rule 482*. Contingent on the presence of an endosperm (character 545, state 1).*

548. (DE139) **Perisperm: Absent** (**0**)**; From nucellar ground tissue** (**1**)**; From nucellar epidermis** (**2**)**.**
Rule 483*. Contingent on the presence of an endosperm (character 545, state 1).* All taxa scored as lacking endosperm were scored not applicable.

549. (DE140) **Embryo, relative to seed interior: Minute, less than half the length of the seed** (**0**)**, Large** (**1**)**.**
Rule 484*. Contingent on the presence of a seed (character 513, state 1).*
Gymnosperms were scored according to the plates of Martin (1946). All other taxa were scored unknown.

550. MOD (DE141) **Cotyledons: Two** (**0**)**; One** (**1**)**;** **Multiple** (**2**)**.**
Rule 485*. Contingent on the presence of a seed (character 513, state 1).*

551. (HB101) **Seeds shed with well-developed embryo: Absent** (**0**)**; Present** (**1**)**.**
Rule 486*. Contingent on the presence of a seed (character 513, state 1).*
Following Hilton and Bateman (2006), all angiosperms were scored present.

552. (HB102) **Seed germination: Hypogeal** (**0**)**; Epigeal** (**1**)**.**
Rule 487*. Contingent on the presence of a seed (character 513, state 1).*
All angiosperms were scored according to DE142. *Oryza* was scored as hypogeal (Tillich 2007).

553 - 562. ***Biochemical characters***

553. (M42) **Lactate fermentation: Absent** (**0**)**; Present** (**1**)**.**

554. (M44) **Hydrogenase produced by vegetative cells: No** (**0**)**; Yes** (**1**)**.**

555. (M45) **Secondary carotenoids: No** (**0**)**; Yes** (**1**)**.**

556. (M46) **Siphonoxanthin: Absent** (**0**)**; Present** (**1**)**.**

557. (M48) **Photosystem II light harvesting complex: low molecular weight** (**0**)**; high molecular weight** (**1**)**.**

558. (M62) **Urea amidolyase produced: No** (**0**)**; Yes** (**1**)**.**

559. (M73) **MOD Glycollate oxidase: Dehydrogenase** (**0**)**; Oxidase** (**1**)**.**

560. (M84) **Monoterpenes: No** (**0**)**; Yes** (**1**)**.**

561. (M85) **Lunularic acid: No** (**0**)**; Yes** (**1**)**.**

562. (M88) **D-Methionine distinguished: No** (**0**)**; Yes** (**1**)**.**

References

Axsmith, B.J., and B.F. Jacobs. 2005. 'The conifer *frenelopsis ramosissima* (cheirolepidiaceae) in the lower cretaceous of texas: Systematic, biogeographical, and paleoecological implications', *International Journal of Plant Sciences*, 166: 327-37.

Barthelemy, D., and Y. Caraglio. 2007. 'Plant architecture: A dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny', *Annals of Botany*, 99: 375-407.

Bechtel, D.B., and Y. Pomeranz. 1977. 'Ultrastructure of mature ungerminated rice (*oryza sativa*) - caryopsis coat and aleurone cells.', *American Journal of Botany*, 64: 966-73.

Beck, C.B. 2010. *An introduction to plant structure and development: Plant anatomy for the twenty-first century* (Cambridge University Press).

Behnke, H.D. 1974. 'Sieve-element plastids of gymnospermae - their ultrastructure in relation to systematics', *Plant Systematics and Evolution*, 123: 1-12.

Bell, P.R., and A.R. Hemsley. 2000. *Green plants: Their origin and diversity* (Cambridge University Press).

Bhatnagar, S.P., and A. Moitra. 1996. *Gymnosperms* (New Age International).

Bierhorst, D.W. 1971. *Morphology of vascular plants* (Macmillan: New York).

Bierhorst, D.W., and Zamora. 1965. 'Primary xylem elements and element associations of angiosperms', *American Journal of Botany*, 52: 657-710.

Biswas, C., and B.M. Johnri. 1997. *The gymnosperms* (Springer-Verlag: Berlin).

Blackmore, S. 1990. 'Sporoderm homologies and morphogenesis in land plants, with a discussion of *echinops sphaerocephala* (compositae)', *Plant Systematics and Evolution, Suppl 5: Morphology, Development, and Systematic Relevance of Pollen and Spores*: 1-12.

Brown, R.C., and B.E. Lemmon. 1990. 'Monoplastidic cell division in lower land plants', *Botanical Journal of Botany*, 77: 559-71.

———. 1992. 'Polar organizers in monoplastidic mitosis of hepatics', *Cell Motility and the Cytoskeleton*, 22: 72-77.

———. 2011a. 'Dividing without centrioles: Innovative plant microtubule organizing centres organize mitotic spindles in bryophytes, the earliest extant lineages of land plants', *AoB Plants*, 2011.

———. 2011b. 'Spores before sporophytes: Hypothesizing the origin of sporogenesis at the algal-plant transition', *New Phytologist*, 190: 875-81.
———. 2013. 'Sporogenesis in bryophytes: Patterns and diversity in meiosis', *Botanical Review*, 79: 178-280.
Brown, R.C., B.E. Lemmon, and O.A. Olsen. 1996. 'Development of the endosperm in rice (*oryza sativa* l): Cellularization', *Journal of Plant Research*, 109: 301-13.
Bruni, A., G. Dallolio, and B. Tosi. 1982. 'A study of the development of raphide-forming cells in *musa paradisiaca* using fluorescence microscopy', *New Phytologist*, 92: 581-87.
Carlquist, S. 1984a. 'Wood anatomy of trimeniaceae', *Plant Systematics and Evolution*, 144: 103-18.
———. 1984b. 'Wood and stem anatomy of lardizabalaceae, with comments on the vining habit, ecology and systematics', *Botanical Journal of the Linnean Society*, 88: 257-77.
———. 1989. 'Wood and bark anatomy of *degeneria*', *Aliso*, 12: 485-95.
———. 1990. 'Wood anatomy of *ascarina* (chloranthaceae) and the tracheid-vessel element transition.', *Aliso*, 12: 667-84.
———. 1992. 'Wood, bark and pith anatomy of old world species of *ephedra* and summary for the genus', *Aliso*, 13: 255-95.
———. 1996a. 'Wood and stem anatomy of menispermaceae', *Aliso*, 14: 155-70.
———. 1996b. 'Wood, bark and stem anatomy of new world species of *gnetum*', *Botanical Journal of the Linnean Society*, 120: 1-19.
———. 1999. 'Wood and bark anatomy of schisandraceae: Implications for phylogeny, habit, and vessel evolution.', *Aliso*, 18: 45-55.
———. 2001. *Comparative wood anatomy: Systematics, ecological, and evolutionary aspects of dicotyledon wood* (Springer-Verlag Berlin Heidelberg: Berlin).
Carlquist, S., and S. Zona. 1988. 'Wood anatomy of papaveraceae, with comments on vessel restriction patterns', *Iawa Bulletin*, 9: 253-67.
Carlquist, S., and E.L. Schneider. 2001. 'Vessels in ferns: Structural, ecological, and evolutionary significance', *American Journal of Botany*, 88: 1-13.
———. 2007. 'Tracheary elements in ferns: New techniques, observations, and concepts', *American Fern Journal*, 97: 199-211.
Chamberlin, M.A., H.T. Horner, and R.G. Palmer. 1994. 'Early endosperm, embryo, and ovule development in *glycine max* (l) merr', *International Journal of Plant Sciences*, 155: 421-36.
Chapman, R.L., and B.H. Good. 1978. 'Ultrastructure of plasmodesmata and cross walls in *cephaleuros*, *phycopeltis* and *trentepohlia* (chroolepidaceae; chlorophyta).', *British Phycological Journal*, 13: 241-46.
Chapman, R.L., and Borkhsenious. 2001. 'Phragmoplast-mediated cytokinesis in *trentepohlia*: Results of tem and immunofluorescence cytochemistry', *International Journal of Systematic and Evolutionary Microbiology*, 51: 759-65.
Chaturvedi, M., K. Datta, and P.K.K. Nair. 1998. 'Pollen morphology of *oryza* (poaceae)', *Grana*, 37: 79-86.
Christiansen, L.C., F. Dal Degan, P. Ulvskov, and B. Borkhardt. 2002. 'Examination of the dehiscence zone in soybean pods and isolation of a dehiscence-related endopolygalacturonase gene', *Plant Cell and Environment*, 25: 479-90.
Church, A.H. 1914. 'On the floral mechanism of *welwitschia mirabilis* (hooker)', *Philosophical Transactions of the Royal Society of London Series B-Containing Papers of a Biological Character*, 205: 47.
Ciampolini, F., K.R. Shivanna, and M. Cresti. 2001. 'Organization of the stigma and transmitting tissue of rice, *oryza sativa* (l.)', *Plant Biology*, 3: 149-55.
Cook, M.E., L.E. Graham, C.E.J. Botha, and C.A. Lavin. 1997. 'Comparative ultrastructure of plasmodesmata of *chara* and selected bryophytes: Toward an elucidation of the evolutionary origin of plant plasmodesmata', *American Journal of Botany*, 84: 1169-78.
Coulter, J.M., and C.J. Chamberlain. 1903. *Morphology of angiosperms*.

Crandall-Stotler, B., R.E. Stotler, and D.G. Long. 2009. 'Phylogeny and classification of the marchantiophyta', *Edinburgh Journal of Botany*, 66: 155-98.

Crane, P.R. 1985. 'Phylogenetic analysis of seed plants and the origin of angiosperms', *Annals of the Missouri Botanical Garden*, 72: 716-93.

Crozier, T.S., and J.F. Thomas. 1993. 'Normal floral ontogeny and cool temperature-induced aberrant floral development in *glycine max* (fabaceae).', *American Journal of Botany*, 80: 429-48.

Dahlgren, R.M.T., H.T. Clifford, and P.F. Yeo. 1985. *The families of the monocotyledons* (Springer-Verlag: Berlin).

Decombeix, A.L., E.L. Taylor, and T.N. Taylor. 2009. 'Secondary growth in *vertebraria* roots from the late permian of antarctica: A change in developmental timing.', *International Journal of Plant Sciences*, 170: 644-56.

Desiro, A., J.G. Duckett, S. Pressel, J.C. Villarreal, and M.I. Bidartondo. 2013. 'Fungal symbioses in hornworts: A chequered history', *Proceedings of the Royal Society B-Biological Sciences*, 280: 8.

DiMichele, W.A., and R.M. Bateman. 1996. 'The rhizomorphic lycopsids: A case study in paleobotanical classification', *Systematic Botany*, 21: 535-52.

Domozych, D.S., and C.E. Domozych. 2014. 'Multicellularity in green algae: Upsizing in a walled complex', *Frontiers in Plant Science*, 5: 649.

Domozych, D.S., M. Ciancia, J.U. Fangel, M.D. Mikkelsen, P. Ulvskov, and W.G.T. Willats. 2012. 'The cell walls of green algae: A journey through evolution and diversity', *Frontiers in Plant Science*, 3: 1-7.

Doyle, J.A. 1996. 'Seed plant phylogeny and the relationships of gnetales', *International Journal of Plant Sciences*, 157: S3-S39.

———. 2006. 'Seed ferns and the origin of angiosperms', *The Journal of the Torrey Botanical Society*, 133: 169-209.

———. 2013. 'Phylogenetic analyses and morphological innovations in land plants', *Annual Plant Reviews*, 45: 1-50.

Doyle, J.A., and P.K. Endress. 2010. 'Integrating early cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots', *Journal of Systematics and Evolution*, 48: 1-35.

Doyle, M.H., and J. Doyle. 1948. 'Pith structure in conifers. 1. Taxodiaceae', *R. Irish Acad. Dublin, Proc.*, 52: 15-39.

Edwards, D. 1994. 'Towards an understanding of pattern and process in the growth of early vascular plants.' in D. S. Ingram and A. Hudson (eds.), *Shape and form in plants and fungi* (Academic Press).

———. 2004. 'Embryophytic sporophytes in the rhynie and windyfield cherts', *Transactions of the Royal Society of Edinburgh*, 94: 397-410.

Edwards, D., L. Axe, and J.G. Duckett. 2003. 'Diversity in conducting cells in early land plants and comparisons with extant bryophytes', *Botanical Journal of the Linnean Society*, 141: 297-347.

Endress, P.K. 2005. 'Carpels in *brasenia* (cabombaceae) are completely ascidiate despite a long stigmatic crest', *Annals of Botany*, 96: 209-15.

———. 2011. 'Angiosperm ovules: Diversity, development, evolution', *Annals of Botany*, 107: 1465-89.

Endress, P.K., and L.D. Hufford. 1989. 'The diversity of stamen structures and dehiscence patterns among magnoliidae.', *Botanical Journal of the Linnean Society*, 100: 45-85.

Esau, K. 1977. *Anatomy of seed plants* (John Wiley and Sons: New York).

Evert, R.F. 2006. *Esau's plant anatomy* (John Wiley and Sons).

Fahn. 1990. *Plant anatomy* (Pergamon Press: New York).

Fernando, D.D., C.R. Quinn, E.D. Brenner, and J.N. Owens. 2010. 'Male gametophyte development and evolution in extant gymnosperms', *International Journal of Plant Reproductive Biology*, 4: 47-63.

Fowke, L.C., and J.D. Pickett-Heaps. 1969. 'Cell division in *spirogyra*. I. Mitosis', *Journal of Phycology*, 5: 240-59.

Friedman, W.E., and J.H. Williams. 2004. 'Developmental evolution of the sexual process in ancient flowering plant lineages', *Plant Cell*, 16: S119-S32.

Friedman, W.E., E.N. Madrid, and J.H. Williams. 2008. 'Origin of the fittest and survival of the fittest: Relating female gametophyte development to endosperm genetics', *International Journal of Plant Sciences*, 169: 79-92.

Friis, E.M., P.R. Crane, and K.R. Pedersen. 2011. *Early flowers and angiosperm evolution* (Cambridge University Press).

Garbary, D.J., K.S. Renzaglia, and J.G. Duckett. 1993. 'The phylogeny of land plants - a cladistic analysis based on male gametogenesis', *Plant Systematics and Evolution*, 188: 237-69.

Gensel, P.G. 1992. 'Phylogenetic relationships of the zosterophylls and lycopsids: Evidence from morphology, paleoecology and cladistic methods of inference.', *Annals of the Missouri Botanical Garden*, 79: 450-73.

Gensel, P.G., and D. Edwards. 2001. *Plants invade the land: Evolutionary and environmental perspectives* (Columbia University Press: New York).

Gerrienne, P. 1988. 'Early devonian plant remains from marchin (north of dinant synclinorium, belgium). 1. *Zosterophyllum deciduum* sp. Nov.', *Review of Palaeobotany and Palynology*, 55: 317-35.

Gerrienne, P., D.L. Dilcher, S. Bergamaschi, I. Milagres, E. Pereira, and M.A.C. Rodrigues. 2006. 'An exceptional specimen of the early land plant *cooksonia paranensis*, and a hypothesis on the life cycle of the earliest eutracheophytes', *Review of Palaeobotany and Palynology*, 142: 123-30.

Ghimire, B., C. Lee, and K. Heo. 2014. 'Leaf anatomy and its implications for phylogenetic relationships in taxaceae s. L', *Journal of Plant Research*, 127: 373-88.

Gifford, E.M., and A.S. Foster. 1989. *Morphology and evolution of vascular plants* (W. H. Freeman and Co.: San Francisco).

Gillespie, L.J., and J.W. Nowicke. 1994. 'Systematic implications of pollen morphology in *gnetum*.', *Acta Botanica Gallica*, 141: 131-39.

Glime, J.M. 2007. 'Bryophyte ecology', EBook sponsored by Michigan Technological University and the International Association of Bryologists.

Gluck, H. 1906. 'Biologische und morphologische untersuchungen über wasser - und sumpfgewachse', *Zweiter Teil* 2: 191-99.

Goffinet, B., and A.J. Shaw. 2009. *Bryophyte biology* (Cambridge University Press).

Goffinet, B., and W.R. Buck. 2013. *The evolution of body form in bryophytes* (John Wiley & Sons, Ltd: Chichester, West Sussex, UK).

Goffinet, B., W.R. Buck, and A.J. Shaw. 2009. 'Morphology, anatomy, and classification of the bryophyta.' in B. Goffinet and A. J. Shaw (eds.), *Bryophyte biology* (Cambridge University Press: Cambridge).

Graham, L.E. 1982. 'The occurrence, evolution, and phylogenetic significance of parenchyma in *coleochaete* breb (chlorophyta)', *American Journal of Botany*, 69: 447-54.

Graham, L.E., and L.W. Wilcox. 2000. *Algae* (Prentice-Hall: NJ).

Halarewicz, A. 2008. 'Morphological and anatomical modification of bracken fern (*pteridium aquilinum* (l.) kuhn.) on serpentinite mable', *EJPAU* 11: 24.

Hao, S.G., and C.B. Beck. 1991. '*Catenalis digitata*, gen. Et sp. Nov., a plant from the lower devonian (siegenian) of yunnan, china', *Canadian Journal of Botany*, 69: 873-82.

Hao, S.G., and J.Z. Xue. 2013. *The early devonian posongchong flora of yunnan: A contribution to an understanding of the evolution and early diversification of vascular plants* (Science Press: Beijing).

Hao, S.G., J.Z. Xue, D.L. Guo, and D.M. Wang. 2010. 'Earliest rooting system and root : Shoot ratio from a new *zosterophyllum* plant', *New Phytologist*, 185: 217-25.

Harris, T.M. 1940. '*Caytonia*', *Annals of Botany*, 4: 713-U3.

Hashimoto, H. 1992. 'Involvement of actin filaments in chloroplast division of the alga *closterium ehrenbergii*.', *Protoplasma*, 167: 88-96.

Hauke, R.L. 1987. 'The ochreole of *equisetum* - a prophyllar shealth.', *American Fern Journal*, 77: 115-23.

Healy, R.A., H.T. Horner, T.B. Bailey, and R.G. Palmer. 2005. 'A microscopic study of trichomes on gynoecia of normal and tetraploid clark cultivars of *glycine max* and seven near-isogenic lines', *International Journal of Plant Sciences*, 166: 415-25.

Heng-Chang, W., J.Q. Li, and A.C. He. 2007. 'Irregular meiotic behavior in *isoetes sinensis* (isoetaceae), a rare and endangered fern in china', *Caryologia*, 60: 358-63.

Hernandez-Castillo, G.R., G.W. Rothwell, and G. Mapes. 2001. 'Thucydiaceae fam. Nov., with a review and reevaluation of paleozoic walchian conifers', *International Journal of Plant Sciences*, 162: 1155-85.

Hickey, L.J., and J.A. Wolfe. 1975. 'Bases of angiosperm phylogeny - vegetative morphology.', *Annals of the Missouri Botanical Garden*, 62: 538-89.

Hilton, J., and R.M. Bateman. 2006. 'Pteridosperms are the backbone of seed-plant phylogeny', *The Journal of the Torrey Botanical Society*, 133: 119-68.

Hodson, M.J., and Bryant. 2012. *Functional biology of plants* (Wiley-Blackwell).

Horner, H.T., R.A. Healy, T. Cervantes-Martinez, and R.G. Palmer. 2003. 'Floral nectary fine structure and development in *glycine max* l. (fabaceae)', *International Journal of Plant Sciences*, 164: 675-90.

Hueber, F.M. 1992. 'Thoughts on the early lycopsids and zosterophylls', *Annals of the Missouri Botanical Garden*, 79: 474-99.

Jarmolenko, A.V. 1939. 'Sur les bois mesozoiques de l'urss depourvus de vaisseaux', *Sovetsk. Bot.*, 6-7: 234-45.

Kedrov, G.B., and A.C. Timonin. 2013. 'Scalariform tracheids in secondary xylem of woody dicotyledons: Distribution, function and evolutionary significance', *Wulfenia*, 20: 43-54.

Kennell, J.C., and H.T. Horner. 1985. 'Megasporogenesis and megagametogenesis in soybean, *glycine max.*', *American Journal of Botany*, 72: 1553-64.

Kenrick, P., and P.R. Crane. 1997. *The origin and early diversification of land plants: A cladistic study* (Smithsonian Institution Press).

Kerp, J.H.F. 1988. 'Aspects of permian paleobotany and palynology. 10. The west european and central european species of the genus *autunia* krasser emend. Kerp (peltaspermaceae) and the form-genus *rhachiphyllum* kerp (callipterid-foliage).', *Review of Palaeobotany and Palynology*, 54: 249-360.

Kidston, R., and W.H. Lang. 1917. 'On old red sandstone plants showing structure, from the rhynie chert bed, aberdeenshire. Part i. *Rhynia gwynne-vaughanii* kidston & lang', *Transactions of the Royal Society of Edinburgh*, 51: 761-84.

———. 1920. 'On old red sandstone plants showing structure from the rhynie chert bed, aberdeenshire. Part ii. Additional notes on *rhynia gwynne-vaughanii* kidston and lang; with descriptions of *rhynia major*, n. Sp. And *hornea lignieri* n. G., n. Sp', *Transactions of the Royal Society of Edinburgh*, 52: 603-27.

King, A.L. 1944. 'The spore discharge mechanism of common ferns', *Proceedings of the National Academy of Sciences of the United States of America*, 30: 155-61.

Klavins, S.D., and L.C. Matten. 1996. 'Reconstruction of the frond of *laceya hibernica*, a lyginopterid pteridosperm from the uppermost devonian of ireland', *Review of Palaeobotany and Palynology*, 93: 253-68.

Knight, C.A., R.B. Clancy, L. Gotzenberger, L. Dann, and J.M. Beaulieu. 2010. 'On the relationship between pollen size and genome size', *Journal of Botany*, Article ID 612017: 7.

Koti, S., K.R. Reddy, V.G. Kakani, D. Zhao, and V.R. Reddy. 2004. 'Soybean (*glycine max*) pollen germination characteristics, flower and pollen morphology in response to enhanced ultraviolet-b radiation', *Annals of Botany*, 94: 855-64.

Kozlowski, T.T. 1972. *Seed biology* (Elsevier).

Kramer, K.U., and P.S. Green. 1990. 'Pteridophytes and gymnosperms.' in Kubitzki K. (ed.), *The families and genera of vascular plants* (Springer: Berlin).

Kuang, A., C.M. Peterson, and R.R. Dute. 1991. 'Pedical abscission and rachis morphology of soybean as influenced by benzylaminopurine and the presence of pods.', *Plant Growth Regulation*, 10: 291-303.

Larsen, J.A. 1927. 'Relation or leaf structure of conifers to light and moisture', *Ecology*, 8: 371-77.

Leavitt, R.G. 1904. 'Trichomes of the root in vascular cryptograms and angiosperms.', *Proceedings of the Boston Society of Natural History* 31: 273-313.

Leng, Q., and E.M. Friis. 2006. 'Angiosperm leaves associated with *sinocarpus infructescences* from the yixian formation (mid-early cretaceous) of ne china', *Plant Systematics and Evolution*, 262: 173-87.

Lersten, N.R., A.R. Czlapinski, J.D. Curtis, R. Freckmann, and H.T. Horner. 2006. 'Oil bodies in leaf mesophyll cells of angiosperms: Overview and a selected survey', *American Journal of Botany*, 93: 1731-39.

Lewis, L.A., and R.M. McCourt. 2004. 'Green algae and the origin of land plants', *American Journal of Botany*, 91: 1535-56.

Ligrone, R., and J.G. Duckett. 1994. 'Thallus differentiation in the marchantialean liverwort *asterella wilmsii* (steph) with particularly reference to lognitudinal arrays of endoplasmic microtubules in the inner cells', *Annals of Botany*, 73: 577-86.

———. 1998. 'The leafy stems of *sphagnum* (bryophyta) contain hightly differentiated polarized cells with axial arrays of endoplasmic microtubules', *New Phytologist*, 140: 567-79.

Ligrone, R., J.G. Duckett, and K. Renzaglia. 2000. 'Conducting tissues and phyletic relationships of bryophytes', *Philosophical Transactions of the Royal Society of London B*, 355: 795-813.

———. 2012. 'Major transitions in the evolution of early land plants: A bryological perspective', *Annals of Botany*.

Ligrone, R., A. Carafa, E. Lumini, V. Bianciotto, P. Bonfante, and J.G. Duckett. 2007. 'Glomeromycotean associations in liverworts: A molecular cellular and taxonomic analysis', *American Journal of Botany*, 94: 1756-77.

Lloyd, R.M., and E.J. Klekowski. 1970. 'Spore germination and viability in pteridophyta: Evolutionary significance of chlorophyllous spores', *Biotropica*, 2: 129-37.

Loconte, H., and D.W. Stevenson. 1990. 'Cladistics of the spermatophyta.', *Brittonia*, 42: 197-211.

Lommasson, R.C., and C.H. Young. 1971. 'Vascularization of fern leaves', *American Fern Society*, 61: 87 - 93.

Manning, F.L. 1914. 'Life history of *porella platyphylla*', *Botanical Gazette*, 57: 320-23.

Marchant, H.J., and J.D. Pickett-Heaps. 1973. 'Mitosis and cytokinesis in *coleochaete scutata*', *Journal of Phycology*, 9: 461 – 71.

Martin, A.C. 1946. 'The comparative internal morphology of seeds.', *American Midland Naturalist*, 36: 513-660.

Melkonian, M. 1982. 'The functional analysis of the flagellar apparatus in green algae.', *Symposia of the Society for Experimental Biology*: 589-606.

Melville, R. 1969. 'Leaf venation patterns and the origin of the angiosperms', *Nature*, 224: 121-25.

Millay, M.A., and T.N. Taylor. 1974. 'Morphological studies of paleozoic saccate pollen.', *Palaeontographica B*, 147: 75 - 99.

Miller, S.S., L.A.A. Bowman, M. Gijzen, and B.L.A. Miki. 1999. 'Early development of the seed coat of soybean (*glycine max*)', *Annals of Botany*, 84: 297-304.

Mishler, B.D., L.A. Lewis, M.A. Buchheim, K.S. Renzaglia, D.J. Garbary, C.F. Delwiche, F.W. Zechman, K.T. S., and R.L. Chapman. 1994. 'Phylogenetic relationships of the "green algae" and the "bryophytes"', *Annals of the Missouri Botanical Garden*, 81: 451-83.

Moestrup, O. 1978. 'Phylogenetic validity of flagellar apparatus in green algae and other chlorophyll a and b containing plants.', *Biosystems*, 10: 117-44.

Newton, A.E., C.J. Cox, J.G. Duckett, J.A. Wheeler, B. Goffinet, T.A.J. Hedderson, and B.D. Mishler. 2000. 'Evolution of the major moss lineages: Phylogenetic analyses based on multiple gene sequences and morphology', *The Bryologist*, 103: 187-211.

Niklas, K.J., and U. Kutschera. 2010. 'The evolution of the land plant life cycle', *Tansley Review, New Phytologist*, 185: 27-41.

Nixon, K.C., W.L. Crepet, D. Stevenson, and E.M. Friis. 1994. 'A re-evaulation of seed plants phylogeny.', *Annals of the Missouri Botanical Garden*, 81: 484-533.

Noblin, X., N.O. Rojas, J. Westbrook, C. Llorens, M. Argentina, and J. Dumais. 2012. 'The fern sporangium: A unique catapult.', *Science*, 335: 1322-22.

Norstog, K.J., and T.J. Nichols. 1997. *The biology of cycads* (Cornell University Press: Ithaca).

Ogura, Y. 1972. 'Comparative anatomy of vegetative organis of the pteridophytes.' in W. Zimmermann, S. Carlquist, P. G. Ozenda and H. D. Wulff (eds.), *Handbuch der pfanzenanatomie*.

Patel, R.N. 1973. 'Wood anatomy of the dicotyledons indigenous to new zealand 3. Monimiaceae and atherospermataceae', *New Zealand Journal of Botany*, 11: 587 - 98.

———. 1992. 'Wood anatomy of the dicotyledons indigenous to new zealand. 22. Proteaceae', *New Zealand Journal of Botany*, 30: 415-28.

Pickett-Heaps, J.D. 1968. 'Ultrastructure and differentiation in *chara* (*fibrosa*). 4. Spermatogenesis.', *Australian Journal of Biological Sciences*, 21: 655.

Pickett-Heaps, J.D. 1975. *Green algae: Structure, reproduction, and evolution in selected genera* (Sinauer Associates).

Pickett-Heaps, J.D., and R. Wetherbee. 1987. 'Spindle function in the green alga *mougeotia* - absence of anaphase a correlates with postmitotic nuclear migration', *Cell Motility and the Cytoskeleton*, 7: 68-77.

Pickett-Heaps, J.D., B.E.S. Gunning, R.C. Brown, B.E. Lemmon, and A.L. Cleary. 1999. 'The cytoplast concept in dividing plant cells: Cytoplasmic domains and the evolution of spatially organized cell division', *American Journal of Botany*, 86: 153-72.

Pigg, K.B., and S. McLoughlin. 1997. 'Anatomically preserved *glossopteris* leaves from the bowen and sydney basins, australia', *Review of Palaeobotany and Palynology*, 97: 339-59.

Posluszny, U., and P.B. Tomlinson. 2003. 'Aspects of inflorescence and floral development in putative basal angiosperm *amborella trichopoda* (amborellaceae)', *Canadian Journal of Botany*, 81: 28-39.

Pressel, S., T. Goral, and J.G. Duckett. 2014. 'Stomatal differentiation and abnormal stomata in hornworts', *Journal of Bryology*, 36: 87-103.

Pressel, S., M.I. Bidartondo, R. Ligrone, and J.G. Duckett. 2010. 'Fungal symbioses in bryophytes: New insights in the twenty first century', *Phytotaxa*, 9: 238-53.

Pryer, K.M. 1999. 'Phylogeny of marsileaceous ferns and relationships of the fossil *hydropteris pinnata* reconsidered', *International Journal of Plant Sciences*, 160: 931-54.

Pryer, K.M., A.R. Smith, and J.E. Skog. 1995. 'Phylogenetic relationships of extant ferns based on evidence from morphology and rbcl sequences', *American Fern Journal*, 85: 205-82.

Raghavan, V. 1989. *Developmental biology of fern gametophytes* (Cambridge University Press: Cambridge, UK).

Rao, A.R., and M. Malaviya. 1964. 'A comparative study of sclereids in some species of *podocarpus*', *Proceedings of the National Institute of Sciences, India*, 31: 67 - 80.

Rashid, A. 1998. *An introduction to bryophyta: Diversity, development and differentiation* (Vikas Publishing House).

Raven, J.A. 1996. 'Into the voids: The distribution, function, development and maintenance of gas spaces in plants', *Annals of Botany*, 78: 137-42.

Ray, M.M., G.W. Rothwell, and R.A. Stockey. 2014. 'Anatomically preserved early cretaceous bennettitalean leaves: *Nilssoniopteris corrugata* n. Sp. From vancouver island, canada.', *Journal of Paleontology*, 88: 1085-93.

Raymond, A., M. Wehner, and S.H. Costanza. 2014. 'Permineralized *alethopteris ambigua* (lesquereux) white: A medullosan with relatively long-lived leaves, adapted for sunny habitats in mires and floodplains', *Review of Palaeobotany and Palynology*, 200: 82-96.

Remy, W., P.G. Gensel, and H. Hass. 1993. 'The gametophyte generation of some early devonian land plants. ', *International Journal of Plant Sciences*, 154: 35-58.

Renzaglia, K.S., and D.J. Garbary. 2001. 'Motile gametes of land plants: Diversity, development, and evolution', *Critical Reviews in Plant Sciences*, 20: 107-213.

Renzaglia, K.S., J.C. Villarreal, and R.J. Duff. 2009. 'New insights into morphology, anatomy, and systematics of hornworts.' in B. Goffinet and A. J. Shaw (eds.), *Bryophyte biology* (Cambridge University Press).

Renzaglia, K.S., R.J. Duff, D.L. Nikrent, and D.J. Garbary. 2000. 'Vegetative and reproductive innovations of early land plants: Implications for a unified phylogeny', *Philosophical Transactions of the Royal Society of London B*, 355: 769-93.

Renzaglia, K.S., R.C. Brown, B.E. Lemmon, J.G. Duckett, and R. Ligrone. 1994. 'Occurrence and phylogenetic significant of monoplastidic meiosis in liverworts', *Canadian Journal of Botany*, 72: 65-72.

Rothwell, G.W. 1975. 'The callistophytaceae (pteridospermopsida): I. Vegetative structures.', *Palaeontographica B*, 151: 171 - 96.

———. 1999. 'Fossils and ferns in the resolution of land plant phylogeny', *Botanical Review*, 65: 188-218.

Rothwell, G.W., and D.M. Erwin. 1985. 'The rhizomorph apex of *paurodendron* - implications for homologies among the rooting organis of lycopsida.', *American Journal of Botany*, 72: 86-98.

Rothwell, G.W., and Karrfalt. 2008. 'Growth, development, and systematics of ferms: Does *botrychium* s.L. (ophiogloassales) really produce secondary xylem?', *American Journal of Botany*, 95: 414-23.

Rudall, P.J., and E.V.W. Knowles. 2013. 'Ultrastructure of stomatal development in early-divergent angiosperms reveals contrasting patterning and pre-patterning', *Annals of Botany*: 1-13.

Rudall, P.J., J. Hilton, and R.M. Bateman. 2013. 'Several developmental and morphogenetic factors govern the evolution of stomatal patterning in land plants', *Tansley Review, New Phytologist*, 200: 598-614.

Rudall, P.J., M.V. Remizowa, A.S. Beer, E. Bradshaw, D.W. Stevenson, T.D. Macfarlane, R.E. Tuckett, S.R. Yadav, and D.D. Sokoloff. 2008. 'Comparative ovule and megagametophyte development in hydatellaceae and water lilies reveal a mosaic of features among the earliest angiosperms', *Annals of Botany*, 101: 941-56.

Ryberg, P.E., E.L. Taylor, and T.N. Taylor. 2012. 'The first permineralized microsporophyll of the glossopteridales: *Eretmonia macloughlinii* sp. Nov.', *International Journal of Plant Sciences*, 173: 812-22.

Rydin, C., and E.M. Friis. 2005. 'Pollen germination in *welwitschia mirabilis* hook. F.: Differences between the polyplicate pollen producing genera of the gnetales', *Grana*, 44: 137-41.

Sawitzky, H., and F. Grolig. 1995. 'Phragmoplast of the green alga *spirogyra* is functionally distinct from the higher plant phragmoplast', *Journal of Cell Biology*, 130: 1359-71.

Schmid, R. 1986. 'On cornerian and other terminology of angiospermous and gymnospermous seed coats - historial perspective and terminological recommendations.', *Taxon*, 35: 476-91.

Schneider, E.L., and S. Carlquist. 2001. 'Sem studies on vessel elements of saururaceae', *Iawa Journal*, 22: 183-92.

Schneider, H. 1996. *Vergleichende wurzelanatomie der farne* (Shaker Verlag).

———. 2013. 'Evolutionary morphology of ferns (monilophytes).' in B. A. Ambrose and M. Purugganan (eds.), *Annual plant reviews: The evolution of plant form* (John Wiley & Sons: Chichester).

Schneider, H., A.R. Smith, and K.M. Pryer. 2009. 'Is morphology really at odds with molecules in estimating fern phylogeny?', *Systematic Biology*, 34: 455-75.

Schofield, W.B. 1985. *Introduction to bryology* (Macmillan: New York).

Seago, J.L., and D.D. Fernando. 2013. 'Anatomical aspects of angiosperm root evolution', *Annals of Botany*, 112: 223-38.

Seago, J.L., L.C. Marsh, K.J. Stevens, A. Soukup, O. Votrubova, and D.E. Enstone. 2005. 'A re-examination of the root cortex in wetland flowering plants with respect to aerenchyma', *Annals of Botany*, 96: 565-79.

Seago, J.L., Jr. 2002. 'The root cortex in the nymphaeaceae, cabombaceae, and nelumbonaceae', *Journal of the Torrey Botanical Society*, 129: 1 - 9.

Serbet, R., and G.W. Rothwell. 1992. 'Characterizing the most primitve seed ferns. 1. A reconstruction of *elkinsia polymorpha.*', *International Journal of Plant Sciences*, 153: 602-21.

Shaw, J., and K. Renzaglia. 2004. 'Phylogeny and diversification of bryophytes', *American Journal of Botany*, 91: 1557-81.

Shimamura, M., Y. Mineyuki, and H. Deguchi. 2003. 'A review of the occurrence of monoplastidic meiosis in liverworts', *Journal of the Hattori Botanical Laboratory*, 94: 179-86.

Shimamura, M., R.C. Brown, B.E. Lemmon, T. Akashi, K. Mizuno, N. Nishihara, K.I. Tomizawa, K. Yoshimoto, H. Deguchi, H. Hosoya, T. Horio, and Y. Mineyuki. 2004. 'Gamma-tubulin in basal land plants: Characterization, localization, and implication in the evolution of acentriolar microtubule organizing centers', *Plant Cell*, 16: 45-59.

Simpson, M.G. 2011. *Plant systematics*.

Singh, V. 2010. *A textbook of botany: Angiosperms* (Rastogi Publications).

Singh, V.P. 2006. *Gymnosperm (naked seeds plant): Structure and development* (Sarup and Sons).

Skorupska, H.T., N.V. Desamero, and R.G. Palmer. 1993. 'Possible homeosis in flower development in cultivated soybean *glycine max* (l.) merr - morphoanatomy and genetics.', *Journal of Heredity*, 84: 97-104.

Sluiman, H.J., F.A.C. Kouwets, and P.C.J. Blommers. 1989. 'Classification and definition of cytokinetic patterns in green algae - sporulation versus (vegetative) cell division', *Archiv Fur Protistenkunde*, 137: 277-90.

Smith, G.M. 1955. *Cryptogamic botany ii. Bryophytes and pteridophytes* (McGraw Hill: New York).

Smith, S.Y., and R.A. Stockey. 2001. 'A new species of *pityostrobus* from the lower cretaceous of california and its bearing on the evolution of pinaceae', *International Journal of Plant Sciences*, 162: 669-81.

Stern, W.L. 1955. 'Xylem anatomy and relationships of gomortegaceae.', *American Journal of Botany*, 42: 874-85.

Stevenson, D.W. 1990. 'Morphology and systematics of cycadales', *Memoirs of the New York Botanical Garden*, 57: 8-55.

———. 2012. 'Gymnosperms.' in B. A. Ambrose and M. Purugganan (eds.), *The evolution of plant form* (Annual Plant Reviews).

Stewart, W.N., and G.W. Rothwell. 2010. *Paleobotany and the evolution of plants* (Cambridge University Press).

Stockey, R.A. 1980. 'Anatomy and morphology of *araucaria sphaerocarpa* carruthers from the jurassic inferior oolite of bruton, somerset.', *Botanical Gazette*, 141: 116-24.

Strullu-Derrien, C., S. McLoughlin, M. Philippe, A. Mork, and D.G. Strullu. 2012. 'Arthropod interactions with bennettitalean roots in a triassic permineralized peat from hopen, svalbard archipelago (arctic)', *Palaeogeography Palaeoclimatology Palaeoecology*, 348: 45-58.

Stubblefield, S.P., and G.W. Rothwell. 1989. '*Cecropsis luculentum* gen. Et sp. Nov. - evidence for heterosporous progymnosperms in the upper pennsylvanian of north america', *American Journal of Botany*, 76: 1415-28.

Suda, S., and Watanabe. 2004. 'Electron microscopy of sexual reproduction in *nephroselmis olivacea* (prasinophyceae, chlorophyta).', *Phycological Research*, 52: 273-83.

Takahashi, C. 1961. 'The growth of protonema cells and rhizoids in bracken', *Cytologia*, 26: 62-66.

Taylor, T.N., E.L. Taylor, and M. Krings. 2009. *Palaeobotany: The biology and evolution of fossil plants. 2nd edition* (Elsevier).

Terrell, E.E., P.M. Peterson, and W.P. Wergin. 2001. 'Epidermal features and spikelet micromorphology in *oryza* and related genera (poaceae: Oryzeae).', *Smithsonian Contributions to Botany*, 91: 1-50.

Thompson, R.H. 1969. 'Sexual reproduction in *chaetosphaeridium globosum* (nordst.) klebahn (chlorophyceae) and description of a species new to science', *Journal of Phycology*, 5: 285-90.

Thompson, W.P., and I.W. Bailey. 1916. 'Are *tetracentron*, *trochodendron* and *drimys* specialized or primitive types?', *Memoirs of the New York Botanical Garden*, 6: 27-32.

Tillich, H.J. 2007. 'Seedling diversity and the homologies of seedling organs in the order poales (monocotyledons)', *Annals of Botany*, 100: 1413-29.

Tilton, V.R., L.W. Wilcox, R.G. Palmer, and M.C. Albertsen. 1984. 'Stigma, style, and obturator of soybean, *glycine max* (l) merr (leguminosae) and their function in the reproductive process.', *American Journal of Botany*, 71: 676-86.

Tomescu, A.M.F., S.E. Wyatt, M. Hasebe, and G.W. Rothwell. 2014. 'Early evolution of the vascular plant body plan - the missing mechanisms', *Current Opinion in Plant Biology*, 17: 126-36.

Tomlinson, P.B., and E.H. Zacharias. 2001. 'Phyllotaxis, phenology and architecture in cephalotaxus, torreya and amentotaxus (coniferales)', *Botanical Journal of the Linnean Society*, 135: 215-28.

Tryon, A.F., and B. Lugardon. 1991. *Spores of the pteridophyta surface wall structure and diversity based on electron microscope studies* (Springer-Verlag: New York).

Vasco, A., R.C. Moran, and B.A. Ambrose. 2013. 'The evolution, morphology, and development of fern leaves', *Frontiers in Plant Science*, 4: 16.

Vaughn, K.C., R. Ligrone, H.A. Owen, J. Hasegawa, E.O. Campbell, K.S. Renzaglia, and J. Mongenajera. 1992. 'The anthocerote chloroplast - a review', *New Phytologist*, 120: 169-90.

Verma, B.K. 2010. *Introduction to taxonomy of angiosperms* (Prentice-Hall of India).

Villarreal, J.C., and K.S. Renzaglia. 2006. 'Sporophyte structure in the neotropical hornwort *phaeomegaceros fimbriatus*: Implications for phylogeny, taxonomy, and character evolution', *International Journal of Plant Sciences*, 167: 413-27.

Villarreal, J.C., and S.S. Renner. 2012. 'Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 18873-78.

Villarreal, J.C., and K.S. Renzaglia. 2015. 'The hornworts: Important advancements in early land plant evolution', *Journal of Bryology*, 37: 157-70.

von Guttenberg, H. 1968a. 'Der primare bau der angiospermenwurzel', *Encyclopedia of Plant Anatomy*, VIII/ 5: 472.

———. 1968b. 'Die wurzel der hydro- und hygrophyten', *Handbuch der Pflanzenanatomie*, VIII: 260 - 303.

Wada, M. 2008. 'Photoresponses in fern gametophytes.' in T. A. RANKER and C. H. HAUFLER (eds.), *The biology and evolution of ferns and lycophytes* (Cambridge Univ. Press).

Wallace, S., A. Fleming, C.H. Wellman, and D.J. Beerling. 2011. 'Evolutionary development of the plant spore and pollen wall', *AoB Plants*.

Wang, X. 2010. *The dawn angiosperms: Uncovering the origin of flowering plants* (Springer).

Warmbrodt, R.D., and R.F. Evert. 1978. 'Comparative leaf structure of 6 species of heterosporous ferns', *Botanical Gazette*, 139: 393-429.

———. 1979a. 'Comparative leaf structure of 6 species of eusporangiate and protoleptosporangiate ferns', *Botanical Gazette*, 140: 153-67.

———. 1979b. 'Comparative leaf structure of several species of homosporous leptosporangiate ferns', *American Journal of Botany*, 66: 412-40.

Watson, J. 1988. 'The cheirolepidiaceae.' in C. B. Beck (ed.), *Origin and evolution of gymnosperms* (Columbia University Press: New York).

Webb, J.A., and P.A. Moore. 1978. *An illustrated guide to pollen analysis*.

Weng, J.K., and C. Chapple. 2010. 'The origin and evolution of lignin biosynthesis', *New Phytologist*, 187: 273-85.

Whittier, D.P., and J.E. Braggins. 2000. 'Observations on the mature gametophyte of *phylloglossum* (lycopodiaceae)', *American Journal of Botany*, 87: 920-24.

Whittier, D.P., and R. Carter. 2007. 'The gametophyte of *lycopodiella prostrata*', *American Fern Journal*, 97: 230-33.

Wilson, C.L. 1979. '*Idiospermum australiense* (*idiospermaceae)* - aspects of vegetative anatomy ', *American Journal of Botany*, 66: 280-89.

Wilson, T.K. 1965. 'The comparative morphology of the canellaceae. Ii. Anatomy of the young stem and node', *American Journal of Botany*, 52: 369 - 78.

Wolniak, S.M., C.M. van der Weele, F. Deeb, T. Boothby, and V.P. Klink. 2011. 'Extremes in rapid cellular morphogenesis: Post-transcriptional regulation of spermatogenesis in *marsilea vestita*', *Protoplasma*, 248: 457-73.

Xue, J.Z., S.G. Hao, and J.F. Basinger. 2010. 'Anatomy of the late devonian *denglongia hubeiensis*, with a discussion of the phyologeny of the cladoxylopsida', *International Journal of Plant Sciences*, 171: 107-20.

Yamaguchi, T., D.Y. Lee, A. Miyao, H. Hirochika, G.H. An, and H.Y. Hirano. 2006. 'Functional diversification of the two c-class mads box genes osmads3 and osmads58 in *oryza sativa*', *Plant Cell*, 18: 15-28.

Zavialova, N., J. van Konijnenburg-van Cittert, and M. Zavada. 2009. 'The pollen ultrastructure of *williamsoniella coronata* thomas (bennettitales) from the bajocian of yorkshire', *International Journal of Plant Sciences*, 170: 1195-200.

Zhang, D.B., and Z.A. Wilson. 2009. 'Stamen specification and anther development in rice', *Chinese Science Bulletin*, 54: 2342-53.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

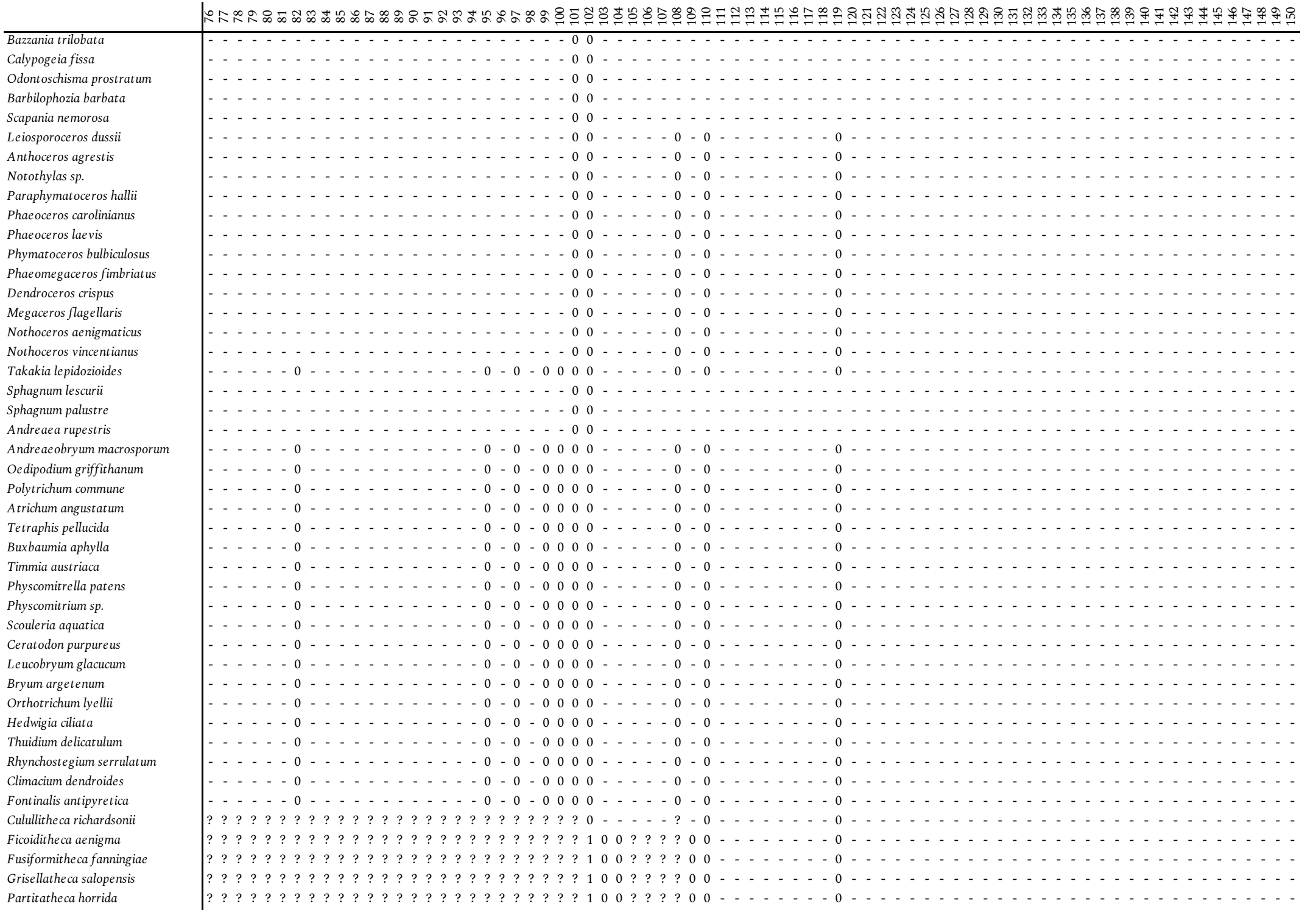
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Sapindopsis	?	1	2	1	?	1	0	1	1	1	0	-	0	?	?	1	0	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1	?	1	1	?	-	?	?	?	?	1	1	?	1	0	?	1	1	0	-	1	?	0	?	1	?	1	?	?	?	?	?	?	?	?	?	?
West bros platanoid	?	1	2	1	?	1	0	1	1	1	0	-	0	?	?	1	0	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1	?	1	1	?	-	?	?	?	?	?	1	?	?	?	?	1	1	0	-	1	?	0	?	1	?	1	?	?	?	?	?	?	?	?	?	?
Spanomera	?	1	2	1	?	1	0	1	1	1	0	-	0	?	?	1	0	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1	?	1	1	?	-	?	?	?	?	?	1	?	?	?	?	1	1	0	-	1	?	0	?	1	?	1	?	?	?	?	?	?	?	?	?	?
Sinocarpus	2	1	2	1	?	1	0	1	1	1	0	-	0	?	?	1	0	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1	?	1	1	?	-	?	1	0	?	?	1	?	?	?	?	1	1	0	-	1	?	0	?	1	?	1	?	?	?	?	?	?	?	?	?	?

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Nephroselmis pyriformis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Characiopodium hindakii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ankistrodesmus falcatus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Draparnaldia mutabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Uronema belkae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mesostigma viride	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlorokybus atmophyticus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Klebsormidium subtile	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Entransia fimbriata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mougeotia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Spirogyra sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zygnema sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zygnemopsis sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mesotaenium endlicherianum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cylindrocystis brebissonii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Closterium lunula	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bambusina borreri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cosmarium granatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cosmarium ochthodes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Staurodesmus convergens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xanthidium antilopaeum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chara vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chaetosphaeridium globosum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Coleochaete scutata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Haplomitrium sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Treubia lacunosa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Blasia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Neohodgsonia mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sphaerocarpos texanus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lunularia cruciata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monoclea gottschei	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Marchantia emarginata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Marchantia polymorpha	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Asterella tenella	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Conocephalum conicum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Riccia austinii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ricciocarpos natans	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pellia neesiana	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pallavicinia lyellii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Riccardiothallus devonicus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Metzgeria crassipilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aneura mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Porella pinnata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Frullania spp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ptilidium pulcherrimum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Schistochila sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-


76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
Bazzania trilobata
Calypogeia fissa
Odontoschisma prostratum
Barbilophozia barbata
Scapania nemorosa
Leiosporoceros dussii
Anthoceros agrestis
Notothylas sp.
Paraphymatoceros hallii
Phaeoceros carolinianus
Phaeoceros laevis
Phymatoceros bulbiculosus
Phaeomegaceros fimbriatus
Dendroceros crispus
Megaceros flagellaris
Nothoceros aenigmaticus
Nothoceros vincentianus
Takakia lepidozioides
Sphagnum lescurii
Sphagnum palustre
Andreaea rupestris
Andreaeobryum macrosporum
Oedipodium griffithanum
Polytrichum commune
Atrichum angustatum
Tetraphis pellucida
Buxbaumia aphylla
Timmia austriaca
Physcomitrella patens
Physcomitrium sp.
Scouleria aquatica
Ceratodon purpureus
Leucobryum glacucum
Bryum argetenum
Orthotrichum lyellii
Hedwigia ciliata
Thuidium delicatulum
Rhynchostegium serrulatum
Climacium dendroides
Fontinalis antipyretica
Culullitheca richardsonii
Ficoiditheca aenigma
Fusiformitheca fanningiae
Grisellatheca salopensis
Partitatheca horrida

| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

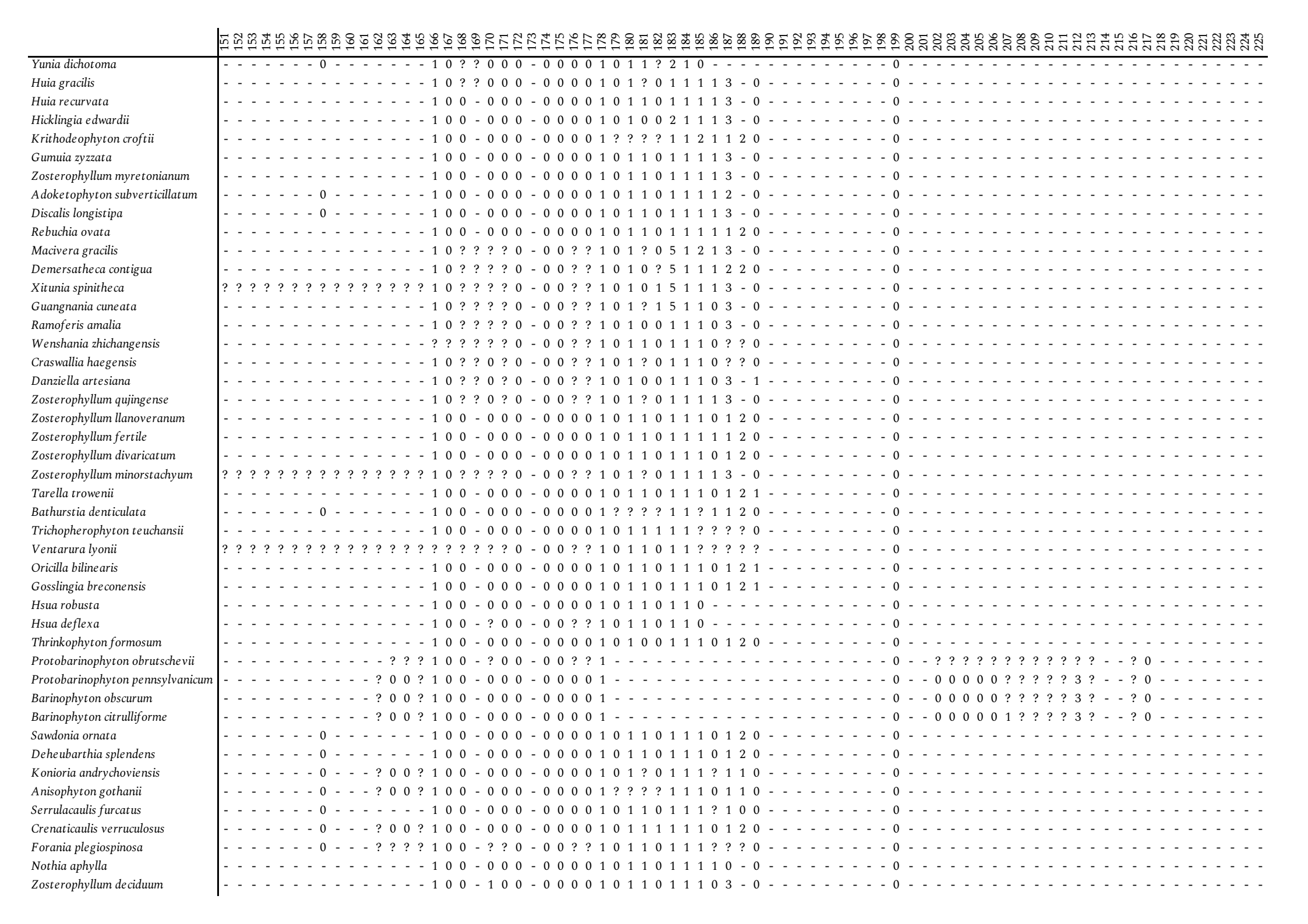
| | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---

	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
Nephroselmis pyriformis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mantoniella squamata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Micromonas pusilla	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tetraselmis carteriiformis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pedinomonas minutissima	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pedinomonas minor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ulva intestinalis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ulva fasciata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ulothrix zonata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cymopolia barbata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Batophora oerstedtii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Codium decorticatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Codium fragile	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cladophoropsis membranacea	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Blastophysa rhizopus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Trentepohlia sp	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cephaleuros parasiticus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cephaleuros virescens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Trebouxia gigantea	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Parietochloris pseudoalveolaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Myrmecia israelensis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Parachlorella kessleri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlorella vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Prototheca wickerhamii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Auxenochlorella protothecoides	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pleurastrum terricola	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlamydopodium vacuolatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dunaliella parva	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dunaliella viridis dumsii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlamydomonas reinhardtii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlamydomonas moewusii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Carteria radiosa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Volvox globator	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Volvox carteri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ettlia minuta	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stephanosphaera pluvialis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gonium pectorale	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mychonastes homosphaera	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Characium perforatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Neochloris aquatica	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Neochloris vigenis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pediastrum duplex	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tetradesmus dimorphus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Scenedesmus obliquus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Scenedesmus fuscus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

| | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 |
|---|---|---

	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
Bazzania trilobata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	-	0	0	0	-	-	0	0	1	1	0	1	-	-	3	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
Partitatheca splendida	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	-	1	0	0	-	0	1	0	-	?	0	0	0	1	0	1	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-


151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
Yunia dichotoma - - - - - - - - 0 - - - - - - - - 1 0 ? ? 0 0 0 - 0 0 0 0 0 1 0 1 1 ? 2 1 0 - - - - - - - - - - - - - - - - 0 - - - - - - - - - - - - - - - -

	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
Baragwanathia longfolia	-	-	-	-	-	-	-	?	?	?	?	?	0	0	?	1	0	0	-	0	0	0	-	0	0	0	0	1	0	1	?	?	1	1	1	0	3	-	0	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-

| | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 |
|---|---|---

| | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 |
|---|---|---

| | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 |
|---|---|---

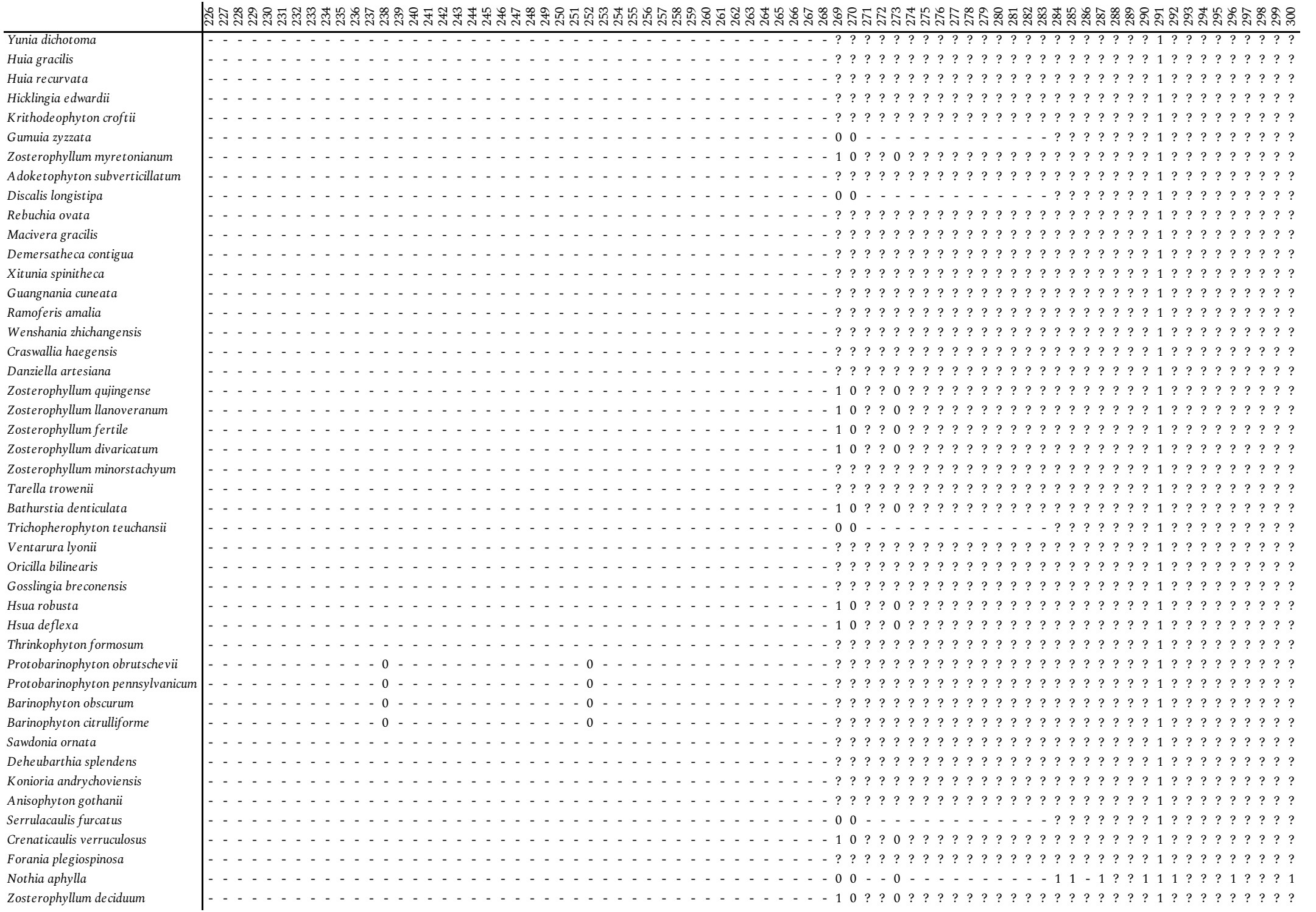
	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
Sapindopsis	0	1	0	?	?	?	?	3	-	-	2	0	2	0	?	1	0	0	-	0	0	0	-	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	0	?	?	1	1	?	1	1	0	1	1	?	1	0	0
West bros platanoid	?	?	?	?	?	?	?	3	-	-	2	0	2	0	?	1	0	0	-	0	0	0	-	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	0	?	?	1	1	?	1	1	0	1	1	?	1	0	0
Spanomera	?	?	?	?	?	?	?	3	-	-	2	0	2	0	?	1	0	0	-	0	0	0	-	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	0	?	?	0	0	?	1	1	0	0	0	0	1	0	0
Sinocarpus	0	1	0	?	?	?	?	3	-	-	?	?	?	0	?	1	0	0	-	0	0	0	-	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	?	0	2	1	3	1	?	0	0	?	?	?	?	?	?	1	?	?	?	0	?	?	?	0

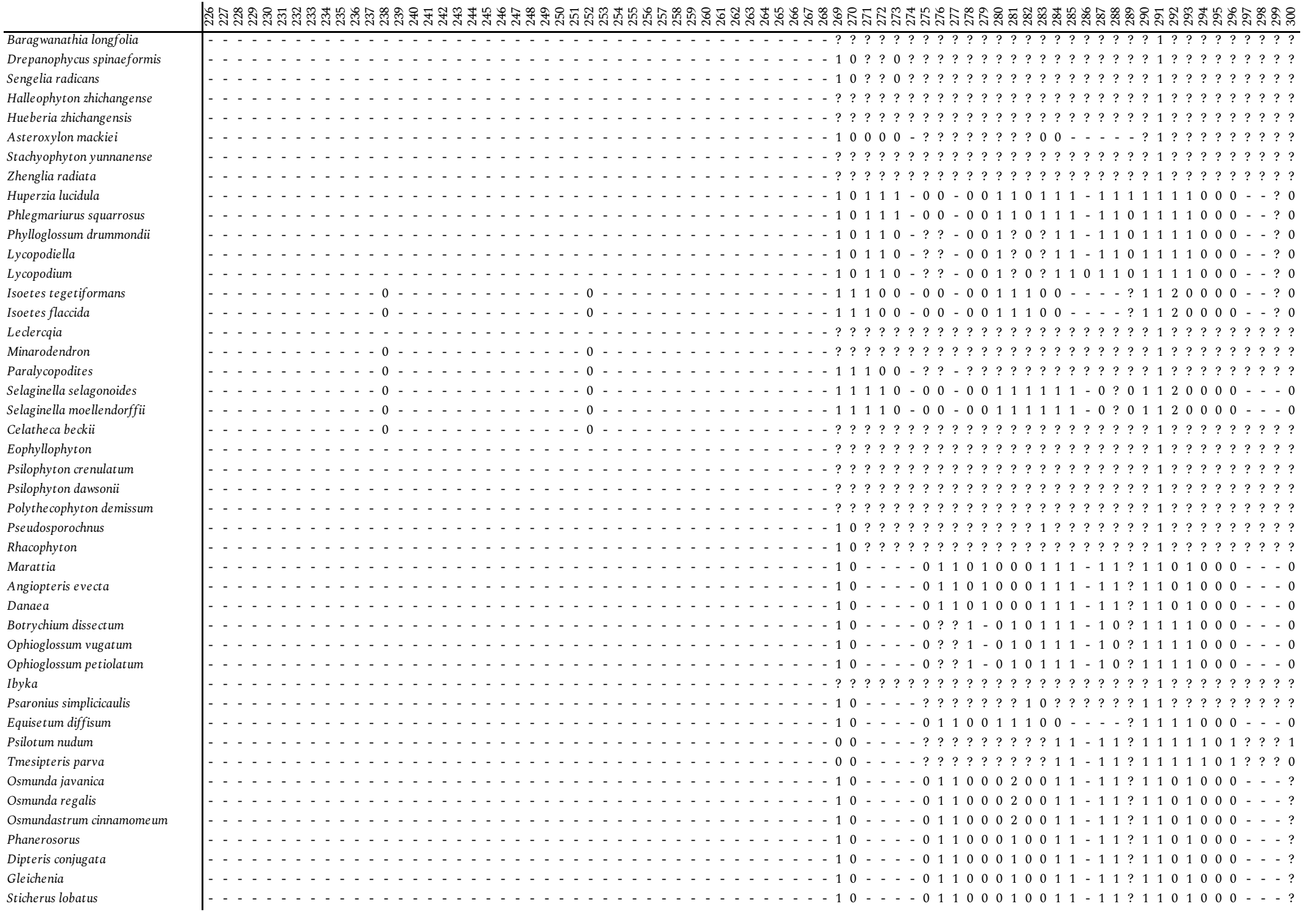
| | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 |
|---|---|---

	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Characiopodium hindakii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Ankistrodesmus falcatus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	?	-	-	-	-	-	-	-	-	-
Draparnaldia mutabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Uronema belkae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Mesostigma viride	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Chlorokybus atmophyticus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Klebsormidium subtile	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Entransia fimbriata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Mougeotia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Spirogyra sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Zygnema sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Zygnemopsis sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Mesotaenium endlicherianum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Cylindrocystis brebissonii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Closterium lunula	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Bambusina borreri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	?	0	0	0	-	-	-	0
Cosmarium granatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Cosmarium ochthodes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Staurodesmus convergens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Xanthidium antilopaeum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-
Chara vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	0	0	0	0	-	-	-	0
Chaetosphaeridium globosum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	?	?	0	0	0	-	-	-	0
Coleochaete scutata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Haplomitrium sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	1	0	2	0	1	0	1	1	1
Treubia lacunosa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	1	0	-	-	-	0
Blasia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Neohodgsonia mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	1	0	-	-	-	0
Sphaerocarpos texanus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Lunularia cruciata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Monoclea gottschei	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Marchantia emarginata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Marchantia polymorpha	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Asterella tenella	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	1
Conocephalum conicum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	1
Riccia austinii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Ricciocarpos natans	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Pellia neesiana	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Pallavicinia lyellii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	1	1	1	0	0
Riccardiothallus devonicus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	?	?	?	?	?	?	?	?
Metzgeria crassipilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Aneura mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	0	0	2	0	0	-	-	-	0
Porella pinnata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	1	0	2	0	0	-	-	-	0
Frullania spp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	1	0	2	0	0	-	-	-	0
Ptilidium pulcherrimum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	1	0	2	0	0	-	-	-	0
Schistochila sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	1	1	1	0	2	0	0	-	-	-	0

| | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 |
|---|---|---

| | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 |
|---|---|---


226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
Yunia dichotoma
Huia gracilis
Huia recurvata
Hicklingia edwardii
Krithodeophyton croftii
Gumuia zyzzata
Zosterophyllum myretonianum
Adoketophyton subverticillatum
Discalis longistipa
Rebuchia ovata
Macivera gracilis
Demersatheca contigua
Xitunia spinitheca
Guangnania cuneata
Ramoferis amalia
Wenshania zhichangensis
Craswallia haegensis
Danziella artesiana
Zosterophyllum qujingense
Zosterophyllum llanoveranum
Zosterophyllum fertile
Zosterophyllum divaricatum
Zosterophyllum minorstachyum
Tarella trowenii
Bathurstia denticulata
Trichopherophyton teuchansii
Ventarura lyonii
Oricilla bilinearis
Gosslingia breconensis
Hsua robusta
Hsua deflexa
Thrinkophyton formosum
Protobarinophyton obrutschevii
Protobarinophyton pennsylvanicum
Barinophyton obscurum
Barinophyton citrulliforme
Sawdonia ornata
Deheubarthia splendens
Konioria andrychoviensis
Anisophyton gothanii
Serrulacaulis furcatus
Crenaticaulis verruculosus
Forania plegiospinosa
Nothia aphylla
Zosterophyllum deciduum


226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
Baragwanathia longfolia - - -

| | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 |
|---|---|---

	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Ginkgo biloba	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	0	2	0	1	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Corystosperms	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Autunia	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Peltaspermum	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Cycas micholitzii	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	0	2	0	1	1	1	-	1	1	0	1	1	2	0	0	0	0	-	-	-	0
Zamiaceae	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	?	2	?	1	1	1	-	1	1	0	1	1	2	0	0	0	0	-	-	-	0
Dioon edule	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	?	2	?	1	1	1	-	1	1	0	1	1	2	0	0	0	0	-	-	-	0
Glossopterids	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	?	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Caytonia	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Bennettitales	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	?	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Pentoxylon	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	2	?	?	?	?	?	?	?	?
Ephedra sinica	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	?	2	?	1	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Welwitschia mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	?	2	?	1	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Gnetum parvifolium	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	0	2	0	1	1	0	?	-	1	?	1	1	2	0	0	0	0	-	-	-	0
Gnetum montanum	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	0	1	1	0	0	0	0	2	0	1	1	0	?	-	1	?	1	1	2	0	0	0	0	-	-	-	0
Amborella trichopoda	0	0	0	0	0	2	?	0	0	0	1	0	1	0	0	0	?	2	0	0	0	0	1	0	0	0	1	1	0	0	0	0	?	0	1	0	0	0	0	0	1	?	1	1	0	-	-	-	0	1	1	0	0	0	?	2	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Hydatellaceae	0	0	?	1	-	-	-	-	-	-	-	0	1	1	1	?	?	?	?	?	?	0	2	0	1	?	1	0	0	0	?	?	?	0	1	0	0	0	0	?	0	?	?	1	0	-	-	-	1	1	1	0	0	0	?	3	1	?	0	-	-	-	-	-	1	1	2	0	0	0	0	-	-	-	0
Cabomba	0	0	0	0	1	1	0	2	?	1	1	0	1	1	1	0	1	0	0	1	0	?	0	2	0	1	0	1	1	0	0	0	0	?	0	1	0	0	2	1	0	0	1	0	1	0	-	-	-	1	1	1	0	0	0	1	3	1	0	0	-	-	-	-	-	1	1	2	0	0	0	0	-	-	-	0
Brasenia	0	0	0	0	1	1	0	2	?	0	0	0	1	0	0	1	0	2	1	0	0	0	2	0	1	0	1	3	0	0	0	0	?	0	1	0	0	1	1	0	0	1	0	1	0	-	-	-	1	1	1	0	0	0	1	3	1	0	0	-	-	-	-	-	1	1	2	0	0	0	0	-	-	-	0
Nuphar advena	0	0	0	0	1	2	0	1	1	0	0	0	1	0	0	1	2	2	1	0	0	0	0	0	2	1	1	2	0	2	0	2	?	0	0	0	0	2	0	0	0	1	1	1	0	-	-	-	1	1	1	0	0	0	1	0	1	1	0	-	-	-	-	?	1	1	2	0	0	0	0	-	-	-	0
Barclaya	0	0	0	0	1	2	2	1	0	0	0	0	1	0	0	1	2	2	?	0	0	0	0	0	0	1	1	2	1	2	0	2	?	0	1	0	0	2	0	0	0	?	?	1	0	-	-	-	1	1	1	0	0	0	1	0	1	?	0	-	-	-	-	-	1	1	2	0	0	0	0	-	-	-	0
Austrobaileya scadens	0	0	0	0	0	2	?	1	0	0	0	0	1	0	0	0	?	2	0	0	1	0	1	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	2	1	0	0	1	1	1	0	-	-	-	0	1	1	0	0	0	0	0	0	0	1	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Trimenia	0	0	?	0	0	2	?	0	0	0	0	0	1	1	0	0	?	1	0	0	0	0	2	0	0	0	1	0	0	0	0	?	0	0	1	0	0	0	0	0	0	1	?	?	1	0	-	-	-	0	1	1	0	0	0	0	0	?	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Kadsura heteroclita	0	0	0	0	0	2	?	1	0	0	0	0	1	1	0	0	?	0	0	1	0	0	1	0	1	0	1	3	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	-	-	-	0	1	1	0	0	0	0	0	?	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Illicium floridanum	0	0	1	0	0	2	?	1	0	0	0	0	1	0	0	0	?	1	0	0	0	0	1	0	1	0	1	2	1	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Illicium parviflorum	0	0	1	0	0	2	?	1	0	0	0	0	1	0	0	0	?	2	0	0	0	0	1	0	1	0	1	2	1	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Ceratophyllum demersum	0	0	?	1	-	-	-	-	-	-	-	0	1	0	1	?	?	?	?	?	?	0	0	0	?	1	1	0	0	0	0	?	?	0	0	0	0	0	0	1	?	0	0	?	0	0	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Hedyosmum	0	0	?	0	1	0	0	0	?	?	0	0	1	1	1	?	?	?	?	?	?	0	0	0	0	1	1	0	0	0	0	?	0	0	0	0	0	0	0	1	0	1	1	?	1	0	-	-	-	0	1	1	0	0	0	0	?	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Ascarina rubricaulis	0	0	?	1	-	-	-	-	-	-	-	0	1	1	0	?	?	0	0	0	?	0	0	0	0	1	1	0	0	0	0	?	0	0	0	0	0	0	0	0	1	1	?	1	0	-	-	-	0	1	1	0	0	0	0	?	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Chloranthus	0	0	?	1	-	-	-	-	-	-	-	0	1	1	?	?	?	?	?	?	?	0	0	0	0	0	1	0	0	0	0	?	1	0	0	0	0	0	0	0	1	0	?	1	0	-	-	-	0	1	1	0	0	0	0	0	0	0	1	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Sarcandra glabra	0	0	?	1	-	-	-	-	-	-	-	0	1	1	1	?	?	?	?	?	?	0	1	0	1	1	1	0	0	0	0	?	1	0	0	0	0	0	0	0	0	?	1	0	-	-	-	0	1	1	0	0	0	0	?	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Piper auritum	0	0	0	1	-	-	-	-	-	-	-	0	1	1	0	1	0	0	0	0	0	0	2	0	1	1	1	1	?	2	1	1	1	0	0	0	0	0	0	?	0	1	?	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Saururus cernuus	0	0	0	1	-	-	-	-	-	-	-	0	1	1	0	1	0	0	1	0	0	0	2	0	1	1	1	1	2	2	1	1	1	0	0	0	0	0	2	1	0	0	1	0	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Lactoris fernandeziana	0	0	0	0	1	0	0	0	?	?	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	2	2	1	0	0	0	0	0	0	2	1	0	0	1	0	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Asaroideae	0	0	0	0	1	0	0	0	1	0	1	0	1	1	0	1	0	1	1	0	0	0	2	0	0	0	1	2	2	2	1	2	0	0	0	0	0	0	2	0	0	0	?	0	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	?	1	?	1	1	2	0	0	0	0	-	-	-	0
Aristolochia elegans	0	0	0	0	1	0	0	1	?	?	1	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	1	2	2	2	1	2	0	0	0	0	0	0	2	0	?	0	1	?	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	?	1	?	1	1	2	0	0	0	0	-	-	-	0
Drimys winteri	0	0	0	0	1	2	1	1	0	0	1	0	1	0	0	1	2	2	1	0	0	0	1	0	1	0	1	1	2	2	1	1	0	0	0	0	0	0	2	0	0	0	1	0	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Canella winterana	0	0	0	0	1	1	0	1	0	0	1	0	1	1	0	1	2	1	0	1	0	0	?	0	1	0	1	1	2	1	0	1	0	0	0	0	0	0	2	0	?	0	1	?	1	0	-	-	-	?	1	1	1	1	1	?	?	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Myristica fragrans	0	0	?	0	1	0	0	0	?	?	1	0	1	1	0	1	0	0	?	1	?	0	?	0	1	0	1	0	1	2	0	?	0	0	0	0	0	0	0	1	0	1	?	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Magnolia grandiflora	0	2	0	0	1	2	0	2	0	0	0	1	1	0	0	0	?	2	0	0	0	0	0	0	0	0	1	1	3	2	2	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Liriodendron	0	2	0	0	1	2	0	1	0	0	0	0	1	1	?	0	0	?	2	0	0	0	0	0	0	0	1	1	1	3	2	2	0	0	0	0	0	0	0	1	1	1	0	1	0	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0
Galbulimima	0	1	0	?	?	?	?	?	?	?	?	1	1	0	0	0	?	2	0	0	1	1	0	0	0	0	1	1	3	2	?	0	0	0	?	?	0	0	0	0	1	0	1	1	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Degeneria	0	1	0	0	0	2	?	1	0	0	1	0	1	0	0	0	?	2	0	0	1	1	0	0	0	0	1	1	0	2	2	1	?	0	0	0	0	0	2	0	0	0	?	?	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Eupomatia bennettii	0	0	0	1	-	-	-	-	-	-	-	1	1	0	0	0	?	2	0	1	1	1	1	0	0	1	1	3	?	2	1	?	0	0	0	0	0	2	0	1	0	?	1	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	?	?	?	?	?	?	1	1	2	0	0	0	0	-	-	-	0
Annona muricata	0	0	0	0	1	2	0	1	0	0	0	0	1	1	0	1	0	2	1	0	0	1	0	0	0	1	1	3	2	1	0	0	0	1	0	0	0	2	1	1	0	1	1	1	0	-	-	-	0	1	1	0	0	0	0	1	0	?	1	1	-	1	1	?	1	1	2	0	0	0	0	-	-	-	0

| | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 |
|---|---|---

| | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 |
|---|---|---

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Nephroselmis pyriformis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	1	1	0	0	0	?	0	0	0	1	-	-
Mantoniella squamata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	?	0	0	0	1	-	-
Micromonas pusilla	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	?	?	0	0	?	?	0	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	1	-	-
Tetraselmis carteriiformis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	?	0	0	0	1	-	-
Pedinomonas minutissima	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	2	0	?	0	1	?	0	1	0	0	0	0	0	0	1	0	0	0	?	0	0	0	1	-	-
Pedinomonas minor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	2	0	?	0	1	?	0	1	0	0	0	0	0	0	1	0	0	0	?	0	0	0	1	-	-
Ulva intestinalis	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	0	0	0	1	1	0	?	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	-	-
Ulva fasciata	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	0	0	0	1	1	0	?	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	-	-
Ulothrix zonata	-	0	-	0	-	0	2	-	-	-	-	0	0	-	?	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1	0	1	-	-
Cymopolia barbata	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	0	1	-	-
Batophora oerstedtii	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	0	0	0	1	1	0	1	1	0	1	0	0	0	1	1	0	0	0	0	1	0	1	-	-
Codium decorticatum	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	0	1	-	-
Codium fragile	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	0	1	-	-
Cladophoropsis membranacea	-	0	-	0	-	0	2	-	-	-	-	0	0	-	0	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	0	0	0	1	?	0	?	1	0	0	1	0	0	?	?	0	0	0	0	1	0	1	-	-
Blastophysa rhizopus	-	0	-	0	-	0	2	-	-	-	-	0	0	-	?	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	0	0	0	1	1	0	?	1	0	?	1	0	0	?	1	?	0	0	0	1	0	1	-	-
Trentepohlia sp	-	0	-	0	-	0	2	-	-	-	-	0	0	-	1	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	?	0	?	1	0	0	0	1	0	1	-	-
Cephaleuros parasiticus	-	0	-	0	-	0	2	-	-	-	-	0	0	-	1	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	?	0	?	1	0	0	0	1	0	1	-	-
Cephaleuros virescens	-	0	-	0	-	0	2	-	-	-	-	0	0	-	1	-	1	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	-	-	-	-	-	-	1	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	?	0	?	1	0	0	0	1	0	1	-	-
Trebouxia gigantea	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0	0	0	?	0	0	0	1	?	0	?	1	0	0	0	0	0	0	1	0	0	?	0	0	0	1	-	-
Parietochloris pseudoalveolaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	0	0	0	1	1	0	?	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	-	-
Myrmecia israelensis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0	0	0	1	0	0	0	1	1	0	?	1	0	0	0	0	0	0	1	0	0	?	0	0	0	1	-	-
Parachlorella kessleri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Chlorella vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Prototheca wickerhamii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Auxenochlorella protothecoides	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Pleurastrum terricola	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0	0	0	?	0	0	0	1	?	0	?	1	0	0	0	0	0	0	1	0	0	?	0	0	0	1	-	-
Chlamydopodium vacuolatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	0	0	0	1	0	0	?	1	0	0	0	0	0	?	1	0	1	0	0	1	0	1	-	-
Dunaliella parva	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	1	?	0	?	0	1	0	0	?	1	0	0	0	0	0	?	1	0	0	0	0	0	1	-	-
Dunaliella viridis dumsii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	1	?	0	?	0	1	0	0	?	1	0	0	0	0	0	?	1	0	0	0	0	0	1	-	-
Chlamydomonas reinhardtii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	?	0	0	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	-	-
Chlamydomonas moewusii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	?	1	0	0	1	0	1	0	1	-	-
Carteria radiosa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	?	1	0	0	1	0	1	0	1	-	-
Volvox globator	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	?	?	0	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
Volvox carteri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	?	?	0	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
Ettlia minuta	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	0	0	0	1	0	0	?	1	0	0	0	0	0	0	1	0	0	?	0	1	0	1	-	-
Stephanosphaera pluvialis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	1	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	1	1	-	-
Gonium pectorale	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	?	2	0	1	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	1	1	-	-
Mychonastes homosphaera	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Characium perforatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	0	0	0	1	1	0	?	1	0	0	0	0	0	?	1	0	1	0	0	1	0	1	-	-
Neochloris aquatica	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	0	0	0	1	0	1	?	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	-	-
Neochloris vigenis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	0	0	0	1	0	1	?	1	0	0	0	0	0	?	1	0	0	0	0	1	0	1	-	-
Pediastrum duplex	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	1	0	0	0	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	-	-
Tetradesmus dimorphus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Scenedesmus obliquus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-
Scenedesmus fuscus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	1	1	1	-	-

| | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 |
|---|---|---

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Bazzania trilobata	-	1	0	1	0	0	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	0	1	1	1	0	1	1	0	-	0	1	3	0	0	0	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	0	0
Calypogeia fissa	-	1	0	1	0	0	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	0	1	1	1	0	1	1	0	-	0	1	3	0	0	0	1	0	0	-	1	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Odontoschisma prostratum	-	1	0	1	0	0	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	0	1	1	1	0	1	1	0	-	0	1	3	0	0	0	1	0	0	-	1	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Barbilophozia barbata	-	1	0	1	0	0	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	0	1	1	1	0	1	1	0	-	0	1	3	0	0	0	1	0	1	-	1	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Scapania nemorosa	-	1	0	1	0	0	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	0	1	1	1	0	1	1	0	-	0	1	3	0	0	0	1	0	1	-	1	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Leiosporoceros dussii	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	0	-	-	-	-	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Anthoceros agrestis	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	2	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Notothylas sp.	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	1	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	0	0
Paraphymatoceros hallii	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	?	?	?	?	?	?	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Phaeoceros carolinianus	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	2	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Phaeoceros laevis	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	2	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	0	0
Phymatoceros bulbiculosus	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	?	?	?	?	?	?	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Phaeomegaceros fimbriatus	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	2	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Dendroceros crispus	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	0	-	-	-	-	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Megaceros flagellaris	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	1	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Nothoceros aenigmaticus	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	?	?	?	?	?	?	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Nothoceros vincentianus	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	1	1	1	1	2	1	0	0	1	1	1	0	1	?	?	0	1	3	0	0	0	1	1	-	2	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Takakia lepidozioides	1	1	?	0	-	0	0	-	-	-	-	0	0	0	1	-	0	0	0	0	1	0	0	2	1	1	1	1	1	0	1	1	?	?	0	1	3	1	1	-	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	1	1
Sphagnum lescurii	1	1	0	0	-	1	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	1	1	1	1	0	1	1	0	-	0	0	3	0	1	-	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	1	1
Sphagnum palustre	1	1	0	0	-	1	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	1	2	1	1	1	1	1	0	1	1	0	-	0	0	3	0	1	-	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	1	1
Andreaea rupestris	1	1	0	0	-	1	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	0	3	1	1	-	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	1	1
Andreaeobryum macrosporum	1	1	0	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	0	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	?	3	1	1	-	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Oedipodium griffithanum	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	1	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Polytrichum commune	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	1	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	1	1
Atrichum angustatum	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Tetraphis pellucida	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Buxbaumia aphylla	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Timmia austriaca	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Physcomitrella patens	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Physcomitrium sp.	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Scouleria aquatica	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Ceratodon purpureus	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Leucobryum glacucum	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Bryum argetenum	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Orthotrichum lyellii	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Hedwigia ciliata	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Thuidium delicatulum	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Rhynchostegium serrulatum	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Climacium dendroides	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Fontinalis antipyretica	1	1	1	0	-	0	0	-	-	-	-	0	0	0	1	-	0	1	1	0	1	0	0	2	1	1	1	1	1	0	1	1	2	2	0	1	3	1	0	1	0	-	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Culullitheca richardsonii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Ficoiditheca aenigma	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Fusiformitheca fanningiae	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Grisellatheca salopensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Partitatheca horrida	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Partitatheca splendida	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Partitatheca cymosa	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Partitatheca sp. murornate	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Partitatheca sp. cheilotetras	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Partitatheca sp. dyadospora	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Lenticulatheca magna	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Lenticulatheca sp.	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sporogonites exuberans	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Tortilicaulis offaeus	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Tortilicaulis transwalliensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Horneophyton lignieri	?	0	?	?	?	0	0	-	-	-	-	?	?	?	1	-	?	?	?	0	1	1	0	?	1	?	?	?	1	0	?	?	?	?	?	?	?	0	0	0	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Caia langii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Aglaophyton major	?	0	?	?	?	0	0	-	-	-	-	?	?	?	?	-	?	?	?	0	?	1	0	?	1	?	?	1	1	0	1	1	?	?	?	?	?	0	0	0	1	1	-	1	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Huvenia kleui	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Stockmansella langii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Rhynia gwynne vaughanii	?	0	?	?	?	0	0	-	-	-	-	?	?	?	?	-	?	?	?	0	?	1	0	?	1	?	?	?	1	0	1	1	4	?	?	?	?	0	0	0	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sennicaulis hippocrepiformis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Taeniocrada dubia	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Catenalis digitata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Salopella marcensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Salopella allenii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Salopella australis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Teruelia diezii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Paracooksonia apiculispora	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Cooksonia cambrensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Cooksonia pertoni	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Cooksonia hemisphaerica	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Cooksonia crassiparietilis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Cooksonia parenensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Concavatheca banksii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Steganotheca striata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Eogaspesia gracilis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Filiformorama simplexa	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Pertonella dactylethra	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Hollandophyton colliculum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Resilitheca salopensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sporathylacium salopense	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Uskiella reticulata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Uskiella spargens	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Tarrantia salopensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Eocooksonia sphaerica	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Aberlemnia caledonica	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Monnowella bennettii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Renalia hueberi	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sartilmania jabachensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Yunia dichotoma	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Huia gracilis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Huia recurvata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Hicklingia edwardii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Krithodeophyton croftii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Gumuia zyzzata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum myretonianum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Adoketophyton subverticillatum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Discalis longistipa	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Rebuchia ovata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Macivera gracilis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Demersatheca contigua	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Xitunia spinitheca	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Guangnania cuneata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Ramoferis amalia	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Wenshania zhichangensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Craswallia haegensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Danziella artesiana	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum qujingense	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum llanoveranum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum fertile	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum divaricatum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum minorstachyum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Tarella trowenii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Bathurstia denticulata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Trichopherophyton teuchansii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Ventarura lyonii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Oricilla bilinearis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Gosslingia breconensis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Hsua robusta	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Hsua deflexa	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Thrinkophyton formosum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Protobarinophyton obrutschevii	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Protobarinophyton pennsylvanicum	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Barinophyton obscurum	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Barinophyton citrulliforme	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sawdonia ornata	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Deheubarthia splendens	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Konioria andrychoviensis	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Anisophyton gothanii	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Serrulacaulis furcatus	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Crenaticaulis verruculosus	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Forania plegiospinosa	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Nothia aphylla	?	0	?	?	?	0	?	-	-	-	-	?	?	?	?	-	?	?	?	0	?	?	?	?	1	?	?	?	1	?	1	?	?	?	?	?	?	0	0	0	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zosterophyllum deciduum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?

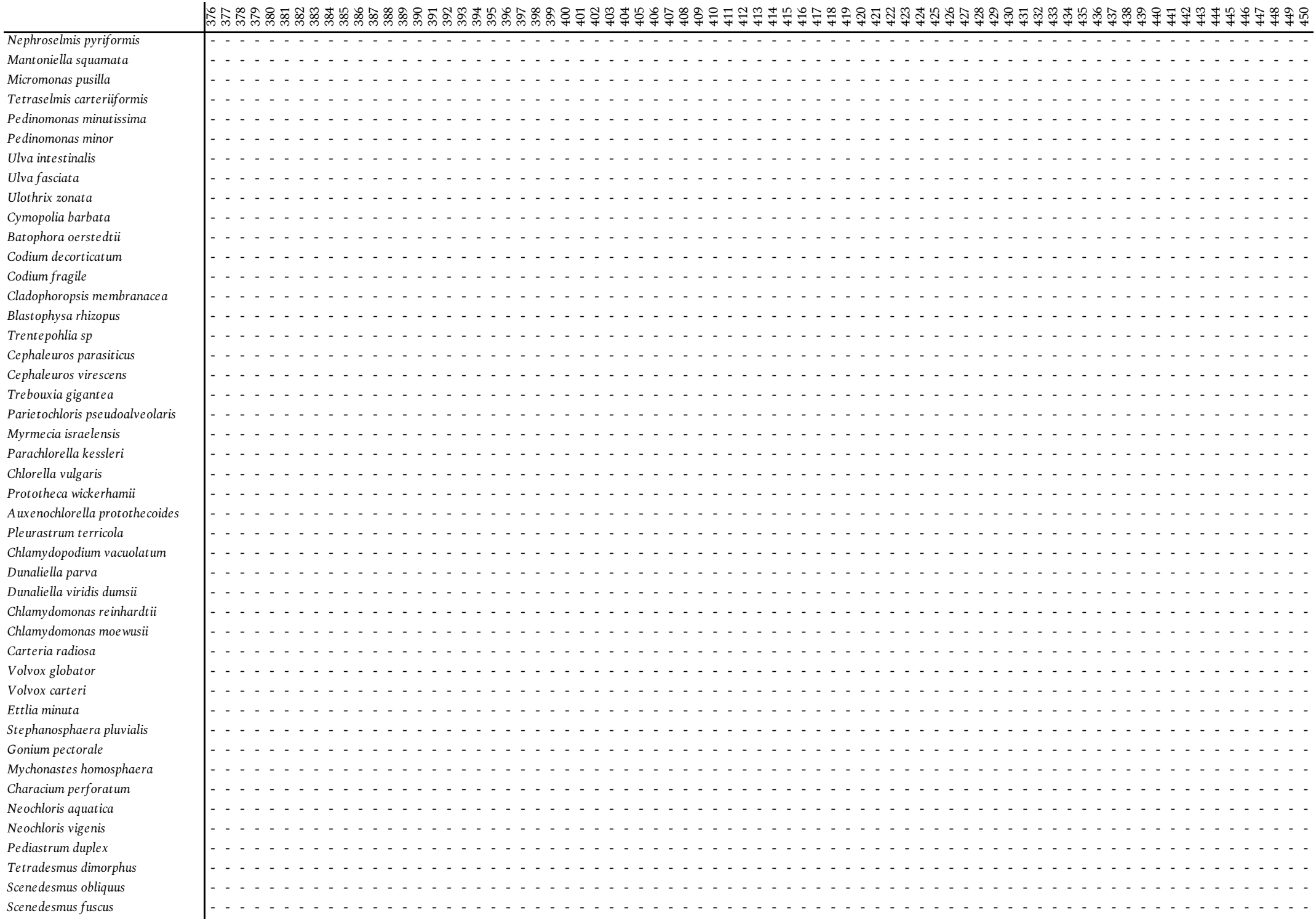
	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Baragwanathia longfolia	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Drepanophycus spinaeformis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sengelia radicans	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Halleophyton zhichangense	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Hueberia zhichangensis	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Asteroxylon mackiei	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Stachyophyton yunnanense	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Zhenglia radiata	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Huperzia lucidula	-	0	0	0	-	0	1	-	-	-	-	0	0	0	1	-	1	-	-	-	1	1	0	0	1	0	-	2	1	0	0	0	0	-	0	?	2	1	0	0	1	1	-	1	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	0
Phlegmariurus squarrosus	-	0	0	0	-	0	1	-	-	-	-	0	0	0	1	-	1	-	-	-	1	1	0	0	1	0	-	2	1	0	0	0	0	-	0	?	2	?	0	0	1	1	-	1	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	?	0	0	0	?	?
Phylloglossum drummondii	-	0	0	0	-	0	0	-	-	-	-	0	0	0	1	-	1	-	-	-	1	1	1	0	1	0	-	2	1	0	0	0	3	1	0	?	2	0	0	0	1	1	-	1	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	0
Lycopodiella	-	0	0	0	-	0	0	-	-	-	-	0	0	0	1	-	1	-	-	-	1	1	1	0	1	0	-	2	1	0	0	0	0	-	0	?	2	1	0	0	1	1	-	1	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Lycopodium	-	0	0	0	-	0	1	-	-	-	-	0	0	0	1	-	1	-	-	-	1	1	1	0	1	0	-	2	1	0	0	0	0	-	0	0	3	0	0	0	1	1	-	0	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	0
Isoetes tegetiformans	-	0	0	0	-	0	0	0	?	?	-	0	0	1	1	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	?	?	?	0	?	5	0	0	0	0	-	-	-	-	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Isoetes flaccida	-	0	0	0	-	0	0	0	?	?	-	0	0	1	1	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	?	?	?	0	?	5	0	0	0	0	-	-	-	-	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Leclercqia	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Minarodendron	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Paralycopodites	?	?	?	?	?	?	0	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Selaginella selagonoides	-	0	0	0	-	0	0	0	?	?	-	0	0	1	1	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	0	?	?	0	1	2	0	0	0	0	-	-	-	-	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Selaginella moellendorffii	-	0	0	0	-	0	0	0	?	?	-	0	0	1	1	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	0	?	?	0	1	2	0	0	0	0	-	-	-	-	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Celatheca beckii	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Eophyllophyton	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Psilophyton crenulatum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Psilophyton dawsonii	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Polythecophyton demissum	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Pseudosporochnus	?	?	?	?	?	?	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Rhacophyton	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Marattia	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	0	3	0	0	?	2	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Angiopteris evecta	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	0	3	?	0	?	2	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Danaea	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	0	0	3	?	0	?	2	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Botrychium dissectum	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	0	0	3	3	0	?	3	0	0	0	1	1	-	1	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Ophioglossum vugatum	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	0	0	3	0	0	?	2	0	0	0	1	1	-	1	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Ophioglossum petiolatum	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	0	0	3	0	0	?	2	0	0	0	1	1	-	1	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Ibyka	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Psaronius simplicicaulis	?	?	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Equisetum diffisum	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	3	3	0	1	2	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	2	1
Psilotum nudum	0	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	1	?	0	?	2	0	0	0	1	1	-	1	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	1
Tmesipteris parva	0	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	1	?	0	?	?	0	0	0	1	1	-	1	0	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	0	?	?
Osmunda javanica	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	3	0	0	0	1	0	0	0	1	1	-	1	1	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	2	1
Osmunda regalis	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	3	0	0	0	1	0	0	0	1	1	-	1	1	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	2	1
Osmundastrum cinnamomeum	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	3	?	0	0	1	0	0	0	1	1	-	1	1	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Phanerosorus	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Dipteris conjugata	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Gleichenia	-	0	1	0	-	0	0	-	-	-	-	1	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Sticherus lobatus	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Hymenophyllum bivalve	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	3	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Crepidomanes venosum	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	3	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Lygodium	-	0	0	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	2	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Dicksonia	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Cyathea	-	0	1	0	-	0	0	-	-	-	-	0	1	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Plagiogyria japonica	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	0	1	1	0	-	2	1	0	1	0	1	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Azolla caroliniana	-	0	?	0	-	0	0	?	?	?	-	0	0	1	0	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	?	0	?	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Marsilea crenata	-	0	?	0	-	0	0	?	?	?	-	0	0	1	0	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	?	0	?	?	0	0	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	2	1
Pilularia americana	-	0	?	0	-	0	0	?	?	?	-	0	0	1	0	0	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	?	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Pteridium	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	2	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	2	1
Adiantum capillus veneris	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	?	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Cheilanthes lindheimeri	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	?	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Pteris vittata	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	?	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Athyrium filix femina	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	?	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Blechnum	-	0	1	0	-	0	0	-	-	-	-	0	0	0	0	-	1	-	-	-	1	1	1	1	1	0	-	2	1	0	1	0	4	?	0	?	1	0	0	0	0	-	-	-	-	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0	0	0	?	?
Pertica	?	0	?	?	?	?	?	-	-	-	-	?	?	?	?	-	?	?	?	?	?

| | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 |
|---|---|---

| | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 |
|---|---|---

	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Sapindopsis	?	0	-	?	?	?	-	?	?	?	?	0	0	1	0	2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
West bros platanoid	?	0	-	?	?	?	-	?	?	?	?	0	0	1	0	2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Spanomera	?	0	-	?	?	?	-	?	?	?	?	0	0	1	0	2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?
Sinocarpus	?	0	-	?	?	?	-	?	?	?	?	0	0	1	0	2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?


376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
Nephroselmis pyriformis
Mantoniella squamata
Micromonas pusilla
Tetraselmis carteriiformis
Pedinomonas minutissima
Pedinomonas minor
Ulva intestinalis
Ulva fasciata
Ulothrix zonata
Cymopolia barbata
Batophora oerstedtii
Codium decorticatum
Codium fragile
Cladophoropsis membranacea
Blastophysa rhizopus
Trentepohlia sp
Cephaleuros parasiticus
Cephaleuros virescens
Trebouxia gigantea
Parietochloris pseudoalveolaris
Myrmecia israelensis
Parachlorella kessleri
Chlorella vulgaris
Prototheca wickerhamii
Auxenochlorella protothecoides
Pleurastrum terricola
Chlamydopodium vacuolatum
Dunaliella parva
Dunaliella viridis dumsii
Chlamydomonas reinhardtii
Chlamydomonas moewusii
Carteria radiosa
Volvox globator
Volvox carteri
Ettlia minuta
Stephanosphaera pluvialis
Gonium pectorale
Mychonastes homosphaera
Characium perforatum
Neochloris aquatica
Neochloris vigenis
Pediastrum duplex
Tetradesmus dimorphus
Scenedesmus obliquus
Scenedesmus fuscus

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

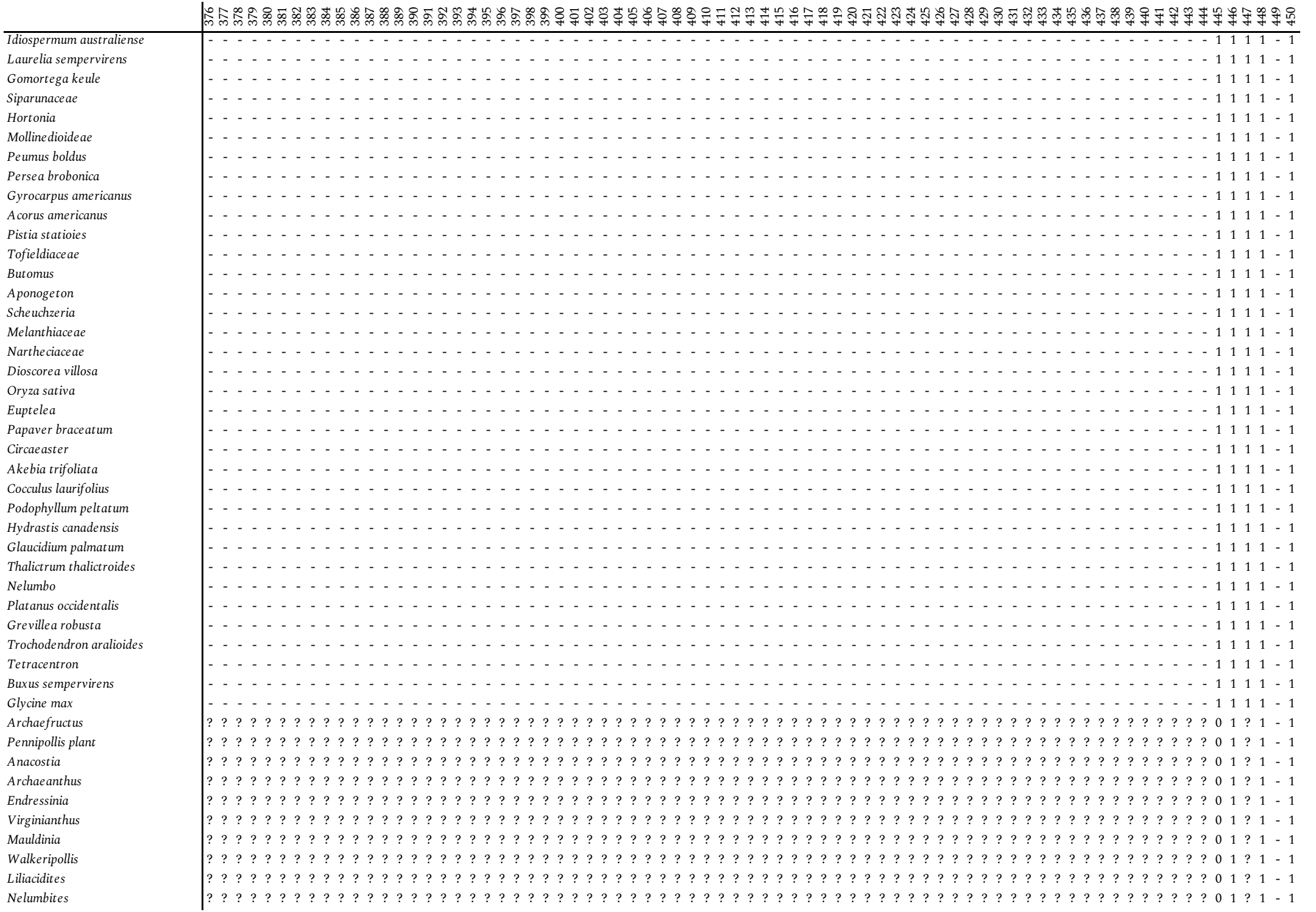
| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---


376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
Idiospermum australiense - - -

| | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 |
|---|---|---

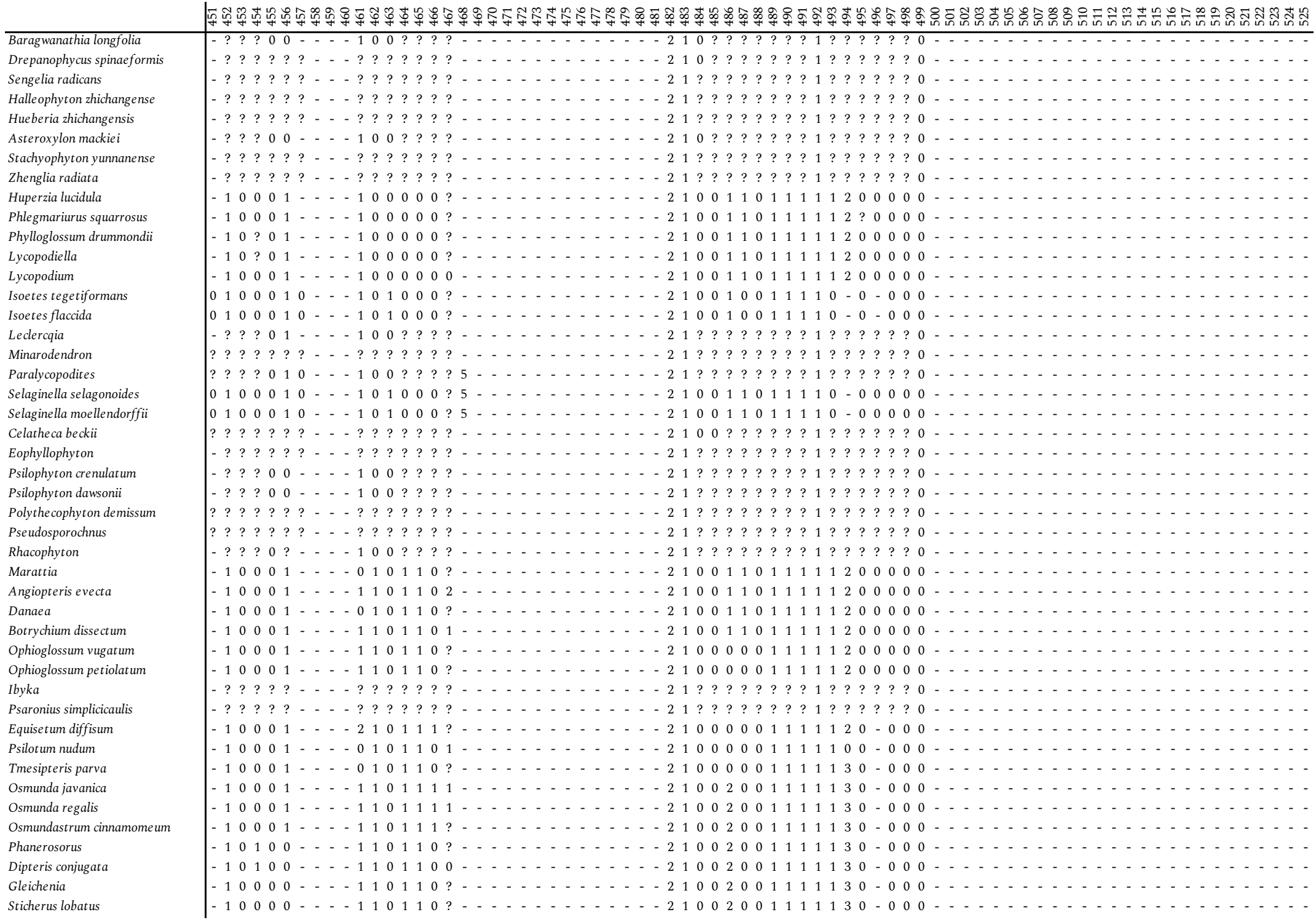
	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Nephroselmis pyriformis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Characiopodium hindakii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ankistrodesmus falcatus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Draparnaldia mutabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Uronema belkae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mesostigma viride	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chlorokybus atmophyticus	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Klebsormidium subtile	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Entransia fimbriata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mougeotia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Spirogyra sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zygnema sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zygnemopsis sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mesotaenium endlicherianum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cylindrocystis brebissonii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Closterium lunula	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bambusina borreri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cosmarium granatum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cosmarium ochthodes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Staurodesmus convergens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xanthidium antilopaeum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	0	0	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chara vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0	-	-	-	-	-	1	1	0	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chaetosphaeridium globosum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0	-	-	-	-	-	1	1	1	0	0	-	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Coleochaete scutata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0	-	-	-	-	-	1	1	1	1	1	1	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Haplomitrium sp.	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Treubia lacunosa	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Blasia sp.	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Neohodgsonia mirabilis	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	?	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sphaerocarpos texanus	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lunularia cruciata	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	?	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monoclea gottschei	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Marchantia emarginata	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Marchantia polymorpha	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	2	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Asterella tenella	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	?	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Conocephalum conicum	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	1	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	?	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Riccia austinii	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	?	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ricciocarpos natans	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	?	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pellia neesiana	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	0	-	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pallavicinia lyellii	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	0	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Riccardiothallus devonicus	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	?	?	?	?	?	?	?	?	1	1	?	?	?	?	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Metzgeria crassipilis	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	0	-	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aneura mirabilis	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	0	-	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Porella pinnata	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	1	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Frullania spp.	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	1	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ptilidium pulcherrimum	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	1	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Schistochila sp.	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	1	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Bazzania trilobata	-	0	-	-	0	1	-	-	-	-	-	0	0	0	0	0	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	0	0	0	0	0	1	1	1	1	1	1	0	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Partitatheca splendida	-	0	-	-	0	1	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Partitatheca cymosa	-	0	-	-	0	1	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Partitatheca sp. murornate	-	0	-	-	0	1	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Partitatheca sp. cheilotetras	-	0	-	-	0	0	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Partitatheca sp. dyadospora	-	0	-	-	0	0	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lenticulatheca magna	-	0	-	-	0	1	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lenticulatheca sp.	-	0	-	-	0	0	-	-	-	-	-	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sporogonites exuberans	-	0	-	-	0	1	-	-	-	-	?	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tortilicaulis offaeus	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tortilicaulis transwalliensis	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Horneophyton lignieri	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Caia langii	-	0	-	-	0	0	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aglaophyton major	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Huvenia kleui	-	0	-	-	0	?	?	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stockmansella langii	-	0	-	-	0	0	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhynia gwynne vaughanii	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sennicaulis hippocrepiformis	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Taeniocrada dubia	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Catenalis digitata	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Salopella marcensis	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Salopella allenii	-	0	-	-	0	0	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Salopella australis	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Teruelia diezii	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Paracooksonia apiculispora	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cooksonia cambrensis	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cooksonia pertoni	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cooksonia hemisphaerica	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cooksonia crassiparietilis	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cooksonia parenensis	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Concavatheca banksii	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Steganotheca striata	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Eogaspesia gracilis	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Filiformorama simplexa	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pertonella dactylethra	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hollandophyton colliculum	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Resilitheca salopensis	-	?	?	?	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sporathylacium salopense	-	?	?	?	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Uskiella reticulata	-	?	?	?	?	?	?	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Uskiella spargens	-	?	?	?	?	?	?	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tarrantia salopensis	-	?	?	?	?	?	?	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Eocooksonia sphaerica	-	?	?	?	?	?	?	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aberlemnia caledonica	-	0	-	-	0	1	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monnowella bennettii	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Renalia hueberi	-	?	?	?	0	0	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sartilmania jabachensis	-	?	?	?	?	?	?	-	-	-	?	?	?	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Yunia dichotoma	-	0	-	-	0	0	-	-	-	-	1	0	0	?	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-



| | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 524 | 525 |
|---|---|---

| | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 524 | 525 |
|---|---|---

	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Idiospermum australiense	?	2	-	-	0	-	0	0	0	?	-	0	0	0	?	0	?	2	-	-	?	0	0	1	1	?	-	0	?	?	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	0	1	2	2	1	?	1	0	0	1	-	1	?	?	?	?	0	1	?
Laurelia sempervirens	4	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	0	?	0	0	0	?	0	0	1	1	4	0	0	2	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	1	1	1	1	1	?	1	0	0	1	-	1	0	0	0	3	0	1	?
Gomortega keule	4	2	-	-	0	-	0	0	1	2	-	0	0	0	0	0	0	?	1	-	-	?	0	0	1	1	3	-	0	-	0	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	?	1	0	0	?	1	1	1	1	?	1	0	0	1	-	1	0	0	0	3	0	0	1
Siparunaceae	4	2	-	-	0	-	0	0	1	2	-	0	0	0	0	0	0	?	1	-	-	?	0	0	2	1	3	-	0	2	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	1	0	?	?	?	?	1	?	1	0	0	1	-	1	0	?	?	?	0	0	1
Hortonia	4	2	-	-	0	-	0	0	1	?	-	0	0	0	?	0	?	1	-	-	?	0	0	1	1	?	-	0	-	0	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	0	1	2	1	?	1	0	0	1	-	1	0	0	0	3	0	0	1
Mollinedioideae	4	2	-	-	0	-	0	0	1	?	-	0	0	0	0	0	?	1	-	-	?	0	1	?	1	3	?	0	-	0	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	1	0	1	2	1	?	1	0	0	1	-	1	0	0	0	3	0	0	1
Peumus boldus	4	2	-	-	0	-	0	0	0	?	-	0	0	0	0	0	?	1	-	-	?	0	0	1	1	3	-	1	-	0	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	1	?	?	1	?	1	0	0	1	-	1	0	0	0	3	0	0	1
Persea brobonica	4	2	-	-	0	-	0	2	0	?	-	0	0	0	?	0	?	1	-	-	?	0	0	1	1	2	?	1	-	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	1	1	0	0	0	?	2	?	1	?	1	0	0	1	-	0	0	0	0	3	0	0	1
Gyrocarpus americanus	4	2	-	-	0	-	0	?	?	?	-	0	0	0	1	0	?	1	-	-	?	0	0	1	1	?	?	1	-	?	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	0	2	2	1	?	1	0	0	1	-	0	0	0	0	3	0	0	1
Acorus americanus	4	2	-	-	0	-	0	0	0	2	-	0	0	0	0	0	?	0	0	0	0	0	0	2	0	4	-	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	?	0	?	1	0	0	?	0	1	0	1	?	1	1	0	1	-	1	0	0	0	0	0	0	0
Pistia statioies	4	2	-	-	0	-	0	0	0	2	-	0	0	0	0	0	?	0	0	0	0	0	0	2	0	4	-	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	?	0	0	1	0	0	?	0	1	0	1	?	1	1	0	1	-	1	0	0	0	0	0	0	0
Tofieldiaceae	4	2	-	-	0	-	0	1	0	1	-	0	0	0	0	0	?	0	0	0	0	0	0	1	0	4	1	0	0	1	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	1	0	1	0	0	0	1	0	0	1	?	1	0	0	1	-	1	1	?	?	0	0	1	?
Butomus	3	2	-	-	0	-	0	1	0	1	-	0	0	0	1	0	?	0	0	0	0	0	0	1	0	4	1	0	0	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	1	2	0	1	0	0	0	0	0	0	1	?	1	0	0	1	-	1	2	?	?	0	0	1	?
Aponogeton	3	2	-	-	0	-	0	1	1	1	-	0	0	0	1	0	?	0	0	0	1	0	0	1	1	2	0	0	0	1	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	0	0	2	0	1	?	1	0	0	1	-	1	3	0	0	0	0	1	?
Scheuchzeria	3	2	-	-	0	-	0	1	0	?	-	0	0	0	1	0	?	1	-	-	?	0	?	1	1	4	?	0	-	?	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	1	0	1	0	0	0	0	2	2	1	?	1	0	0	1	-	1	3	1	0	0	0	1	?
Melanthiaceae	4	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	?	0	0	0	1	0	0	1	0	4	1	0	0	1	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	?	?	1	0	1	?	1	0	0	1	-	1	1	0	0	0	0	1	?
Nartheciaceae	?	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	?	0	0	0	0	0	0	1	0	4	1	0	0	1	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	?	?	1	0	1	?	1	0	0	1	-	1	1	0	0	0	0	1	?
Dioscorea villosa	4	2	-	-	0	-	0	0	0	2	-	0	0	0	0	0	?	0	0	0	0	0	0	1	0	4	-	0	0	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	0	?	0	1	?	1	0	0	1	-	1	0	0	0	?	0	1	?
Oryza sativa	?	2	-	-	0	-	?	0	?	2	-	0	0	0	0	0	?	0	0	?	?	0	0	0	1	?	-	0	?	?	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	?	?	?	1	0	0	?	?	?	?	1	?	1	0	0	1	-	?	0	?	0	?	0	1	?
Euptelea	4	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	0	0	1	1	1	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	0	2	1	1	?	1	0	0	1	-	1	2	0	0	4	0	1	?
Papaver braceatum	4	2	-	-	0	-	0	?	1	2	-	0	0	0	?	0	?	4	-	-	1	0	0	1	1	4	?	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	1	0	0	2	1	?	1	0	0	1	-	1	1	?	?	?	0	1	?
Circaeaster	4	2	-	-	0	-	1	0	1	?	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	-	0	?	?	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	?	1	1	0	?	?	?	?	1	?	1	1	0	1	-	1	0	?	?	?	0	1	?
Akebia trifoliata	4	2	-	-	0	-	0	0	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	-	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	1	0	1	0	0	?	0	?	?	1	?	1	0	0	1	-	0	1	0	0	0	0	0	0
Cocculus laurifolius	4	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	2	1	4	0	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	?	0	1	0	1	?	1	0	0	1	-	1	0	0	0	0	?	0	1
Podophyllum peltatum	4	2	-	-	0	-	0	1	0	2	-	0	0	0	?	0	?	4	-	-	1	0	0	1	1	4	2	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	2	0	1	0	0	0	1	2	?	1	?	1	0	0	1	-	?	1	0	0	0	0	0	0
Hydrastis canadensis	4	2	-	-	0	-	1	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	0	0	1	2	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	1	1	2	2	1	?	1	0	0	1	-	1	1	0	0	0	0	1	?
Glaucidium palmatum	?	2	-	-	0	-	0	1	1	2	-	0	0	0	?	0	?	4	-	-	?	0	0	1	1	4	0	0	1	2	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	1	?	1	0	0	?	?	2	2	1	?	1	1	0	1	-	1	1	0	0	0	0	1	?
Thalictrum thalictroides	4	2	-	-	0	-	0	0	1	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	-	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	?	0	1	0	0	0	0	2	?	1	?	1	1	0	1	-	1	1	0	0	0	0	1	?
Nelumbo	3	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	0	1	4	0	0	1	1	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	1	1	0	0	0	1	2	2	1	?	1	0	0	1	-	1	0	0	0	0	0	1	?
Platanus occidentalis	4	2	-	-	0	-	0	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	0	0	1	2	1	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	?	1	0	0	?	1	1	2	1	?	1	0	0	1	-	1	0	1	0	0	0	1	?
Grevillea robusta	4	2	-	-	0	-	0	?	0	1	-	0	0	0	0	0	?	4	-	-	1	0	0	1	2	4	0	0	0	2	0	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	?	1	0	0	?	1	0	?	1	?	1	0	0	1	-	?	0	0	0	0	0	?	1
Trochodendron aralioides	4	2	-	-	0	-	1	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	2	0	1	2	1	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	1	0	2	1	?	1	0	0	1	-	1	2	?	?	4	0	1	?
Tetracentron	4	2	-	-	0	-	1	1	0	2	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	2	0	1	2	1	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	0	0	0	2	1	?	1	0	0	1	-	1	2	?	?	4	0	1	?
Buxus sempervirens	4	2	-	-	0	-	?	1	0	?	-	0	0	0	0	0	?	4	-	-	1	0	0	1	1	4	0	0	?	1	1	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	0	0	0	1	0	0	1	1	2	1	1	?	1	0	0	1	-	1	1	?	?	?	0	?	0
Glycine max	4	2	-	-	0	-	?	?	?	2	-	0	0	0	0	0	?	4	-	-	?	0	0	2	1	?	?	0	?	?	?	2	1	1	0	1	1	0	1	1	1	1	1	2	0	1	1	1	1	?	?	?	?	1	0	0	?	?	?	?	1	?	1	0	0	1	-	?	1	?	0	?	0	1	?
Archaefructus	?	?	?	?	?	-	0	0	0	?	-	?	?	?	?	?	?	0	0	0	1	0	0	1	0	?	?	0	?	?	?	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	0	?	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	1	?	?	?	?	1	?	?
Pennipollis plant	?	?	?	?	?	-	0	1	1	1	-	0	0	?	?	?	?	0	0	0	1	0	0	2	1	2	0	0	0	2	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	?	?	?	0	0	0
Anacostia	?	?	?	?	?	-	0	0	1	?	-	?	?	?	?	?	?	0	0	0	1	0	0	1	1	4	2	0	1	1	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	2	?	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1	0	0	0	0	0	0
Archaeanthus	?	?	?	?	?	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	0	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Endressinia	?	?	?	?	?	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Virginianthus	?	?	?	?	?	-	0	1	0	1	-	0	0	?	?	?	?	0	0	0	1	0	0	2	1	4	0	0	0	2	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Mauldinia	?	?	?	?	?	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	1
Walkeripollis	?	?	?	?	?	-	0	1	1	2	-	0	0	?	?	?	?	0	1	0	1	0	1	1	1	4	0	0	1	1	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Liliacidites	?	?	?	?	?	-	0	1	0	1	-	0	0	?	?	?	?	0	0	0	0	0	0	0	1	0	4	1	0	0	1	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Nelumbites	?	?	?	?	?	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Sapindopsis	?	?	?	?	?	-	0	1	0	2	-	0	0	?	?	?	?	4	-	-	1	0	0	2	1	4	0	0	1	2	?	2	1	?	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	?	?	?	?	1	?
West bros platanoid	?	?	?	?	?	-	0	1	0	2	-	0	0	?	?	?	?	4	-	-	1	0	0	2	1	4	0	0	1	2	?	2	1	?	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Spanomera	?	?	?	?	?	-	1	1	0	2	-	0	0	?	?	?	?	4	-	-	1	0	0	1	1	4	0	0	1	2	?	2	1	?	?	?	?	?	?	?	?	?	1	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	?
Sinocarpus	?	?	?	?	?	-	?	?	?	?	-	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	2	1	?	?	?	?	?	?	?	?	1	?	?	?	?	?	1	?	?	?	?	1	0	?	?	?	?	?	?	1	?	?	?	1	?	?	?	0	?	?	?	?

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Nephroselmis pyriformis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Mantoniella squamata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	0	0	0	0
Micromonas pusilla	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Tetraselmis carteriiformis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	0	0	0	0	0
Pedinomonas minutissima	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Pedinomonas minor	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Ulva intestinalis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	0	0	?	0	0	0
Ulva fasciata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	1	?	0	0	0	?	0	0	0
Ulothrix zonata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	0	?	0	0	0	0
Cymopolia barbata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	0	0	0	0
Batophora oerstedtii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	0	0	0	0
Codium decorticatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	1	1	?	0	0	0	0
Codium fragile	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	1	?	0	0	0	0
Cladophoropsis membranacea	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	0	?	0	0	0	0
Blastophysa rhizopus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	1	1	0	0	0	0	0
Trentepohlia sp	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Cephaleuros parasiticus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	0	0	0	0	0	0
Cephaleuros virescens	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	0	0	0	0	0	0
Trebouxia gigantea	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Parietochloris pseudoalveolaris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	0	0	?	?	?	0	0	0
Myrmecia israelensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Parachlorella kessleri	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	0	?	0	?	0	0	0
Chlorella vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	1	0	0	0	0
Prototheca wickerhamii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Auxenochlorella protothecoides	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	0	0	?	?	?	0	0	0
Pleurastrum terricola	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	0	0	0	0	0
Chlamydopodium vacuolatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Dunaliella parva	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Dunaliella viridis dumsii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0
Chlamydomonas reinhardtii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	1	?	0	0	1	0	0	0	0
Chlamydomonas moewusii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Carteria radiosa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Volvox globator	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Volvox carteri	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Ettlia minuta	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Stephanosphaera pluvialis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	1	?	0	0	0
Gonium pectorale	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Mychonastes homosphaera	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	0	0	0
Characium perforatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Neochloris aquatica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Neochloris vigenis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Pediastrum duplex	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Tetradesmus dimorphus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Scenedesmus obliquus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	1	1	0	?	?	0	0	0	0
Scenedesmus fuscus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	0	0	0	0

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Characiopodium hindakii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Ankistrodesmus falcatus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Draparnaldia mutabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	1	?	0	0	0
Uronema belkae	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	?	0	0	0
Mesostigma viride	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Chlorokybus atmophyticus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Klebsormidium subtile	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Entransia fimbriata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Mougeotia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Spirogyra sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Zygnema sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Zygnemopsis sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Mesotaenium endlicherianum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Cylindrocystis brebissonii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Closterium lunula	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Bambusina borreri	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Cosmarium granatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Cosmarium ochthodes	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Staurodesmus convergens	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Xanthidium antilopaeum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Chara vulgaris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Chaetosphaeridium globosum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Coleochaete scutata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	0
Haplomitrium sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Treubia lacunosa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Blasia sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Neohodgsonia mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Sphaerocarpos texanus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Lunularia cruciata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Monoclea gottschei	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Marchantia emarginata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Marchantia polymorpha	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Asterella tenella	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Conocephalum conicum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Riccia austinii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Ricciocarpos natans	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Pellia neesiana	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Pallavicinia lyellii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Riccardiothallus devonicus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Metzgeria crassipilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Aneura mirabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Porella pinnata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Frullania spp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Ptilidium pulcherrimum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Schistochila sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Bazzania trilobata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Calypogeia fissa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Odontoschisma prostratum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Barbilophozia barbata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	?	0
Scapania nemorosa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	1	1	0
Leiosporoceros dussii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Anthoceros agrestis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Notothylas sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Paraphymatoceros hallii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Phaeoceros carolinianus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Phaeoceros laevis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Phymatoceros bulbiculosus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Phaeomegaceros fimbriatus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Dendroceros crispus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Megaceros flagellaris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Nothoceros aenigmaticus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Nothoceros vincentianus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Takakia lepidozioides	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Sphagnum lescurii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Sphagnum palustre	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Andreaea rupestris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Andreaeobryum macrosporum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Oedipodium griffithanum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Polytrichum commune	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Atrichum angustatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Tetraphis pellucida	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Buxbaumia aphylla	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Timmia austriaca	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Physcomitrella patens	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Physcomitrium sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Scouleria aquatica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Ceratodon purpureus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Leucobryum glacucum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Bryum argetenum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Orthotrichum lyellii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Hedwigia ciliata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Thuidium delicatulum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Rhynchostegium serrulatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Climacium dendroides	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Fontinalis antipyretica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	?	?	1	0	0	1
Culullitheca richardsonii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Ficoiditheca aenigma	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Fusiformitheca fanningiae	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Grisellatheca salopensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Partitatheca horrida	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Partitatheca splendida	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Partitatheca cymosa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Partitatheca sp. murornate	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Partitatheca sp. cheilotetras	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Partitatheca sp. dyadospora	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Lenticulatheca magna	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Lenticulatheca sp.	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Sporogonites exuberans	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Tortilicaulis offaeus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Tortilicaulis transwalliensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Horneophyton lignieri	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Caia langii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Aglaophyton major	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Huvenia kleui	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Stockmansella langii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Rhynia gwynne vaughanii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Sennicaulis hippocrepiformis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Taeniocrada dubia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Catenalis digitata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Salopella marcensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Salopella allenii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Salopella australis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Teruelia diezii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Paracooksonia apiculispora	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Cooksonia cambrensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Cooksonia pertoni	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Cooksonia hemisphaerica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Cooksonia crassiparietilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Cooksonia parenensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Concavatheca banksii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Steganotheca striata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Eogaspesia gracilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Filiformorama simplexa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Pertonella dactylethra	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Hollandophyton colliculum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Resilitheca salopensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Sporathylacium salopense	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Uskiella reticulata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Uskiella spargens	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Tarrantia salopensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Eocooksonia sphaerica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Aberlemnia caledonica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Monnowella bennettii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Renalia hueberi	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Sartilmania jabachensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Yunia dichotoma	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Huia gracilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Huia recurvata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Hicklingia edwardii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Krithodeophyton croftii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Gumuia zyzzata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum myretonianum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Adoketophyton subverticillatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Discalis longistipa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Rebuchia ovata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Macivera gracilis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Demersatheca contigua	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Xitunia spinitheca	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Guangnania cuneata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Ramoferis amalia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Wenshania zhichangensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Craswallia haegensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Danziella artesiana	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum qujingense	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum llanoveranum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum fertile	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum divaricatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum minorstachyum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Tarella trowenii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Bathurstia denticulata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Trichopherophyton teuchansii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Ventarura lyonii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Oricilla bilinearis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Gosslingia breconensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Hsua robusta	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Hsua deflexa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Thrinkophyton formosum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Protobarinophyton obrutschevii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Protobarinophyton pennsylvanicum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Barinophyton obscurum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Barinophyton citrulliforme	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Sawdonia ornata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Deheubarthia splendens	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Konioria andrychoviensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Anisophyton gothanii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Serrulacaulis furcatus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Crenaticaulis verruculosus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Forania plegiospinosa	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Nothia aphylla	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zosterophyllum deciduum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Baragwanathia longfolia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Drepanophycus spinaeformis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Sengelia radicans	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Halleophyton zhichangense	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Hueberia zhichangensis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Asteroxylon mackiei	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Stachyophyton yunnanense	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Zhenglia radiata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Huperzia lucidula	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Phlegmariurus squarrosus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Phylloglossum drummondii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Lycopodiella	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Lycopodium	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Isoetes tegetiformans	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Isoetes flaccida	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Leclercqia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Minarodendron	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Paralycopodites	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Selaginella selagonoides	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Selaginella moellendorffii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Celatheca beckii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Eophyllophyton	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Psilophyton crenulatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Psilophyton dawsonii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Polythecophyton demissum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Pseudosporochnus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Rhacophyton	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Marattia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Angiopteris evecta	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Danaea	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Botrychium dissectum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Ophioglossum vugatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Ophioglossum petiolatum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Ibyka	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Psaronius simplicicaulis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Equisetum diffisum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Psilotum nudum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Tmesipteris parva	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Osmunda javanica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Osmunda regalis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Osmundastrum cinnamomeum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Phanerosorus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Dipteris conjugata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Gleichenia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Sticherus lobatus	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Hymenophyllum bivalve	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Crepidomanes venosum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Lygodium	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Dicksonia	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Cyathea	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Plagiogyria japonica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Azolla caroliniana	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Marsilea crenata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Pilularia americana	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Pteridium	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Adiantum capillus veneris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Cheilanthes lindheimeri	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Pteris vittata	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Athyrium filix femina	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Blechnum	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	0	1	?	1	?	0	1
Pertica	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Tetraxylopteris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Rellimia thomsonii	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Archaeopteris	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Cecropsis	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?	?	?	?	?	?	?	?	?
Elkinsia	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Laceya	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Bilignea	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Lyrasperma	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Heterangium	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Lyginopteris	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Pitus	?	?	?	?	?	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Nystroemia	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Medullosa	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Quaestora	-	-	-	?	?	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Callistophyton	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Cordaixylon	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Mesoxylon	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Shanxioxylon	-	-	-	0	0	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
Sergeia	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Barthelia	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Emporia	-	-	-	0	0	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?
Thucydia	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Cheirolepidiaceae	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Pinus engelmannii	-	-	-	0	0	0	-	-	-	1	2	1	1	?	?	?	0	1	?	1	?	0	1
Podocarpus coriaceus	-	-	-	0	0	0	-	-	-	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Araucariaceae	-	-	-	0	0	0	-	-	-	?	2	1	1	?	?	?	0	1	?	1	?	0	1
Cupressus dupreziana	-	-	-	0	0	0	-	-	-	1	2	1	1	?	?	?	0	1	?	1	?	0	1
Cephalotaxus harringtonia	-	-	-	0	0	0	-	-	-	1	2	1	1	?	?	?	0	1	?	1	?	0	1
Taxus baccata	-	-	-	0	1	0	-	-	-	?	0	1	1	?	?	?	0	1	?	1	?	0	1

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Ginkgo biloba	-	-	-	0	0	0	-	-	-	1	0	1	0	?	?	?	0	1	?	1	?	0	1
Corystosperms	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Autunia	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Peltaspermum	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Cycas micholitzii	-	-	-	0	0	0	-	-	-	?	0	1	0	?	?	?	0	1	?	1	?	0	1
Zamiaceae	-	-	-	0	0	0	-	-	-	1	0	1	?	?	?	?	0	1	?	1	?	0	1
Dioon edule	-	-	-	0	0	0	-	-	-	1	0	1	?	?	?	?	0	1	?	1	?	0	1
Glossopterids	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Caytonia	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Bennettitales	-	-	-	0	0	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?
Pentoxylon	-	-	-	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Ephedra sinica	-	-	-	0	0	0	-	-	-	1	0	1	1	?	?	?	0	1	?	1	?	0	1
Welwitschia mirabilis	-	-	-	0	0	0	-	-	-	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Gnetum parvifolium	-	-	-	0	0	0	-	-	-	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Gnetum montanum	-	-	-	0	0	0	-	-	-	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Amborella trichopoda	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Hydatellaceae	0	0	0	1	0	1	0	0	1	0	?	1	0	?	?	?	0	1	?	1	?	0	1
Cabomba	0	0	0	1	0	1	2	0	1	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Brasenia	0	0	0	1	0	1	2	0	1	0	0	1	0	?	?	?	0	1	?	1	?	0	1
Nuphar advena	0	0	0	1	1	1	0	0	1	0	0	1	0	?	?	?	0	1	?	1	?	0	1
Barclaya	0	0	0	1	0	1	0	0	1	0	0	1	0	?	?	?	0	1	?	1	?	0	1
Austrobaileya scadens	0	0	0	0	0	1	?	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Trimenia	0	0	0	0	0	1	?	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Kadsura heteroclita	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Illicium floridanum	1	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Illicium parviflorum	1	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Ceratophyllum demersum	0	0	?	0	0	1	0	0	0	1	0	1	?	?	?	?	0	1	?	1	?	0	1
Hedyosmum	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Ascarina rubricaulis	0	0	2	0	0	1	?	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Chloranthus	0	0	0	0	0	1	0	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Sarcandra glabra	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Piper auritum	0	0	1	0	0	1	?	0	1	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Saururus cernuus	1	0	1	0	0	1	0	0	1	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Lactoris fernandeziana	1	0	0	0	0	1	0	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Asaroideae	1	0	2	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Aristolochia elegans	?	0	2	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Drimys winteri	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Canella winterana	0	0	0	0	?	1	?	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Myristica fragrans	1	0	2	0	1	1	1	0	0	0	0	1	0	?	?	?	0	1	?	1	?	0	1
Magnolia grandiflora	1	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Liriodendron	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Galbulimima	0	0	0	0	0	1	?	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Degeneria	1	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Eupomatia bennettii	0	0	0	0	0	1	?	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Annona muricata	?	0	0	0	0	1	0	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Idiospermum australiense	0	0	?	0	0	1	?	1	0	1	0	1	0	?	?	?	0	1	?	1	?	0	1
Laurelia sempervirens	0	0	0	0	0	1	?	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Gomortega keule	0	0	0	0	0	1	0	0	0	1	0	1	?	?	?	?	0	1	?	1	?	0	1
Siparunaceae	0	0	?	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Hortonia	0	0	0	0	0	1	?	0	0	0	0	1	?	?	?	?	0	1	?	1	?	0	1
Mollinedioideae	0	0	0	0	0	1	?	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Peumus boldus	0	0	0	0	0	1	0	0	0	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Persea brobonica	0	0	0	0	0	1	1	1	0	1	0	1	0	?	?	?	0	1	?	1	?	0	1
Gyrocarpus americanus	0	0	0	0	0	1	0	1	0	1	0	1	0	?	?	?	0	1	?	1	?	0	1
Acorus americanus	0	0	1	0	0	1	0	0	2	0	1	1	1	?	?	?	0	1	?	1	?	0	1
Pistia statioies	0	0	1	0	0	1	0	0	2	0	1	1	1	?	?	?	0	1	?	1	?	0	1
Tofieldiaceae	1	0	0	0	0	1	2	0	0	1	1	1	1	?	?	?	0	1	?	1	?	0	1
Butomus	1	0	0	0	0	1	2	1	0	1	1	1	1	?	?	?	0	1	?	1	?	0	1
Aponogeton	?	0	0	0	0	1	2	1	0	1	1	1	1	?	?	?	0	1	?	1	?	0	1
Scheuchzeria	1	0	0	0	0	1	2	?	0	1	1	1	0	?	?	?	0	1	?	1	?	0	1
Melanthiaceae	1	0	0	0	0	1	2	0	0	0	1	1	1	?	?	?	0	1	?	1	?	0	1
Nartheciaceae	1	0	0	0	0	1	2	0	0	0	1	1	1	?	?	?	0	1	?	1	?	0	1
Dioscorea villosa	?	0	2	0	0	1	1	0	0	?	1	1	0	?	?	?	0	1	?	1	?	0	1
Oryza sativa	0	0	?	0	0	1	1	0	0	1	1	1	0	?	?	?	0	1	?	1	?	0	1
Euptelea	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Papaver braceatum	2	0	2	0	0	1	1	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Circaeaster	0	1	?	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Akebia trifoliata	0	0	0	0	0	1	0	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Cocculus laurifolius	0	0	0	0	0	1	1	0	0	1	0	1	1	?	?	?	0	1	?	1	?	0	1
Podophyllum peltatum	?	0	0	0	?	1	1	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Hydrastis canadensis	1	0	0	0	0	1	1	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Glaucidium palmatum	1	0	0	0	0	1	1	0	0	1	0	1	1	?	?	?	0	1	?	1	?	0	1
Thalictrum thalictroides	1	0	0	0	0	1	1	0	0	0	0	1	1	?	?	?	0	1	?	1	?	0	1
Nelumbo	0	0	0	0	0	1	1	1	0	1	0	1	0	?	?	?	0	1	?	1	?	0	1
Platanus occidentalis	0	0	0	0	0	1	1	0	0	1	0	1	1	?	?	?	0	1	?	1	?	0	1
Grevillea robusta	0	0	0	0	0	1	1	?	0	1	0	1	1	?	?	?	0	1	?	1	?	0	1
Trochodendron aralioides	1	0	2	0	0	1	0	0	0	0	0	1	0	?	?	?	0	1	?	1	?	0	1
Tetracentron	1	0	2	0	0	1	?	0	0	0	0	1	0	?	?	?	0	1	?	1	?	0	1
Buxus sempervirens	?	0	0	0	0	1	0	0	0	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Glycine max	1	0	?	0	0	1	1	0	0	?	0	1	1	?	?	?	0	1	?	1	?	0	1
Archaefructus	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Pennipollis plant	0	0	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Anacostia	0	0	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Archaeanthus	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Endressinia	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Virginianthus	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Mauldinia	0	0	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Walkeripollis	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Liliacidites	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Nelumbites	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548
Sapindopsis	1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
West bros platanoid	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Spanomera	1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Sinocarpus	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

Appendix 5 – Publications

This appendix contains all papers published by the author during the course of the PhD. They are presented in chronological order as follows:

Schneider, Harald; Liu, Hongmei; Clark, James; Hidalgo, Oriane; Pellicer, Jaume; Zhang, Shouzhou; Kelly, Laura J; Fay, Michael F; Leitch, Ilia J (2015) Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in the royal ferns. *New Phytologist*,207(1):10- 13.

Puttick, Mark N; Clark, James; Donoghue, Philip CJ (2015) Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms *Proc. R. Soc. B*,282:1820

Clark, James; Hidalgo, Oriane; Pellicer, Jaume; Liu, Hongmei; Marquardt, Jeannine; Robert, Yannis; Christenhusz, Maarten; Zhang, Shouzhou; Gibby, Mary; Leitch, Ilia J Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytologist* 210(3):1072-1082

Clark, James W; Donoghue, Philip CJ (2017) Constraining the timing of whole genome duplication in plant evolutionary history. *Proc. R. Soc. B* 284: 1858

Puttick, Mark N; O'Reilly, Joseph E; Tanner, Alastair R; Fleming, James F; Clark, James; Holloway, Lucy; Lozano-Fernandez, Jesus; Parry, Luke A; Tarver, James E; Pisani, Davide (2017) Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proc. R. Soc. B*. 284:1846

Clark, James W; Donoghue, Philip CJ (2018) Whole-Genome Duplication and Plant Macroevolution. *Trends in plant science*

Morris, Jennifer L; Puttick, Mark N; Clark, James W; Edwards, Dianne; Kenrick, Paul; Pressel, Silvia; Wellman, Charles H; Yang, Ziheng; Schneider, Harald; Donoghue, Philip CJ (2018) The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences* 115(10):E2274-E2283

Betts, Holly C; Puttick, Mark N; Clark, James W; Williams, Tom A; Donoghue, Philip CJ; Pisani, Davide (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin *Nature ecology & evolution*

Deline, Bradley; Greenwood, Jennifer M; Clark, James W; Puttick, Mark N; Peterson, Kevin J; Donoghue, Philip CJ (2018) Evolution of metazoan morphological disparity *Proceedings of the National Academy of Sciences*

Letters

Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in the royal ferns

Progress in understanding genome evolution has recovered not only evidence for differences in genome dynamics between the major lineages of land plants but also putative links between genome dynamics, evolvability and the assembly of species diversity (Leitch & Leitch, 2012, 2013). A new, inspiring fossil discovery now provides us with a unique insight into the evolution of genome size in a species belonging to a remarkable group of ancient plants, the royal ferns (Osmundaceae). Based on the analysis of an Early Jurassic fossil with hypothesized close affinities to the extant *Osmundastrum cinnamomeum,* Bomfleur *et al.* (2014) argue that the exceptionally well preserved chromosomes and cell nuclei provide evidence for 'genomic stasis' over *c.* 180 million yr. Their argument coincides with previously reported morphological stasis over > 200 million yr in the same group of fossils (Phipps *et al.*, 1998; Serbet & Rothwell, 1999), and Manton's (1950) suggestion that the geological longevity and morphological stability of Osmundaceae was, in part, due to the stability in chromosome number and structure she observed. Thus the fossil suggests a correlation between genotypic and phenotypic stasis over time in these ancient ferns.

The hypothesis that royal ferns exhibit 'genomic stasis', might appear to be consistent with the conservation in chromosome number and near complete absence of polyploidy (Zhang *et al.*, 2008; see Tsutsumi *et al.*, 2011). Yet 'genomic stasis' implies that the underlying genomic processes have had no impact on genomic content over time, something that cannot be inferred solely from analysing fossils. Certainly, genomic turnover can take place without impacting genome size. For example, *Medicago truncatula* and *Lotus japonicus* are closely related angiosperm species with the same genome size (Cheng & Grant, 1973; Arumuganathan & Earle, 1991) but show many differences in genic content; several thousand gene families are present in one species but not the other (Varshney *et al.*, 2012). Dynamic genome evolution is also reflected in changes in transposable element (TE) content and diversity; *Aegilops cylindrica* and *A. geniculata* are close grass relatives with similar genome sizes, but show contrasting patterns of TE amplification and deletion over time (Senerchia *et al.*, 2013). These examples illustrate how relative stasis in genome size and chromosome number may belie genomic turnover. Clearly, fossil evidence of DNA content alone is not sufficient to conclude that royal fern genomes have become effectively 'fossilized'. Here we explore new and existing genomic data to determine whether there is indeed evidence that royal fern genomes are frozen in time.

Genome size and chromosomes

Bomfleur *et al.* (2014) noted the similarity in nucleus size and hence, by proxy, genome size between the fossil and extant species *Osmundastrum cinnamomeum*. To examine if there is evidence of genome size stasis across the whole family we analysed genome size variation and reconstructed the ancestral genome size (Table 1; Fig. 1). By combining previously reported values with newly obtained data (Table 1), we now have genome sizes for half the extant species diversity in Osmundaceae, including at least one representative from each of the four recognized genera and three subgenera of *Osmunda*. Overall, genome sizes varied 1.34-fold with the smallest genome in *Osmunda claytoniana* (1C = 13.46 pg) and the largest in *Todea barbara* (1C = 19.46 pg) (Fig. 1). Using Bayesian approaches we inferred the ancestral genome size for the crown group of Osmundaceae to be 1C = 15.9 pg, with evidence of both increases (e.g. branch leading to *Todea* and *Leptopteris*) and decreases (e.g. branch leading to *Osmunda* subg. *Osmunda*) in genome size over time (Fig. 1). It is noted that if *Osmundastrum* is placed as sister to *Osmunda*, as suggested by morphological similarities, then the ancestral genome size of the crown group is estimated to be 1C = 17.1 pg. Nevertheless, even with this higher value, increases and decreases in genome size are still evident (data not shown). Overall, we have uncovered evidence for both evolutionary changes in genome size and limited genome size variation among extant species of royal ferns, albeit low compared with some other fern lineages (e.g. 1C-values in Salviniales range 5.29-fold based on just three genome size estimates, and 6.41-fold in Ophioglossales based on just five estimates) and ferns as a whole where genome sizes range 94-fold (based on data for 128 species, see Leitch & Leitch, 2013).

An analysis of available chromosome data also reveals a similar story, with evidence of some, albeit limited, variation in the organization of DNA within the chromosomes, despite all but one of the 14 chromosome counts for Osmundaceae reporting $2n = 44$. For example, recent karyological studies in *Osmunda* have provided evidence of chromosome rearrangements (Zhang *et al.*, 2008), interspecific hybridization and allopolyploidy (Tsutsumi *et al.*, 2011).

Substitution rate

Such rather limited genome size and chromosome diversity coincides with a notably low substitution rate of *c.* 1.1×10^{-4} substitutions per site per million years (8.0×10^{-5} to 1.4×10^{-4})

Table 1 Genome size estimates of royal ferns

Species	1C-value (SD)[a] (pg)	CV[b] (%)	Calibration standard[c]	Buffer[d]	Voucher
Leptopteris hymenophylloides (A. Rich.) C. Presl	18.92 (0.17)	2.75	*Vicia*	1	RBGE 19992140
Osmunda banksiifolia (C. Presl) Kuhn	17.25 (0.23)	2.87	*Pisum*	1	Liu *et al.* OS1
Osmunda claytoniana L. (China)	15.29 (0.12)	3.76	*Allium*	1	K, Fay s.n.
Osmunda claytoniana L. (Canada)	13.46 (0.08)	—[e]			Bainard *et al.* (2011)
Osmunda japonica Thunb.	14.15 (0.06)	2.08	*Pisum*	1	Liu *et al.* OS2
Osmunda regalis L. var. *spectabilis* (Willd.) A.Gray	14.01 (0.19)	—[e]			Bainard *et al.* (2011)
Osmunda vachellii Hook.	16.68 (0.23)	3.51	*Pisum*	1	Liu *et al.* OS3
Osmundastrum cinnamomeum (L.) C.Presl	15.45 (0.08)	—[e]			Bainard *et al.* (2011)
Todea barbara T. Moore	19.48 (0.08)	2.35	*Allium*	2	19652792

Four out of the eight species were measured for the first time, while three values were obtained from Bainard *et al.* (2011). The previous measurement of *Todea barbara* published by Obermayer *et al.* (2002) is replaced by a new, more accurate measurement. Herbarium specimens are cited for new measurements.
[a]Nuclear DNA contents were estimated using flow cytometry as described in Pellicer *et al.* (2014); SD, standard deviation.
[b]Coefficient of variation.
[c]*Allium cepa* 'Alice' 2C = 34.89 pg, *Pisum sativum* 'Ctirad' 2C = 9.09 pg and *Vicia faba* 'Inovec' 2C = 26.90 pg (Doležel *et al.*, 1992, 1998).
[d]Nuclei isolation buffer 1, 'General purpose isolation buffer' (GPB, Loureiro *et al.*, 2007) supplemented with 3% polyvinylpyrrolidone (PVP-40); 2, Ebihara's buffer (Ebihara *et al.*, 2005).
[e]Coefficient of variation not given in Bainard *et al.* (2011).

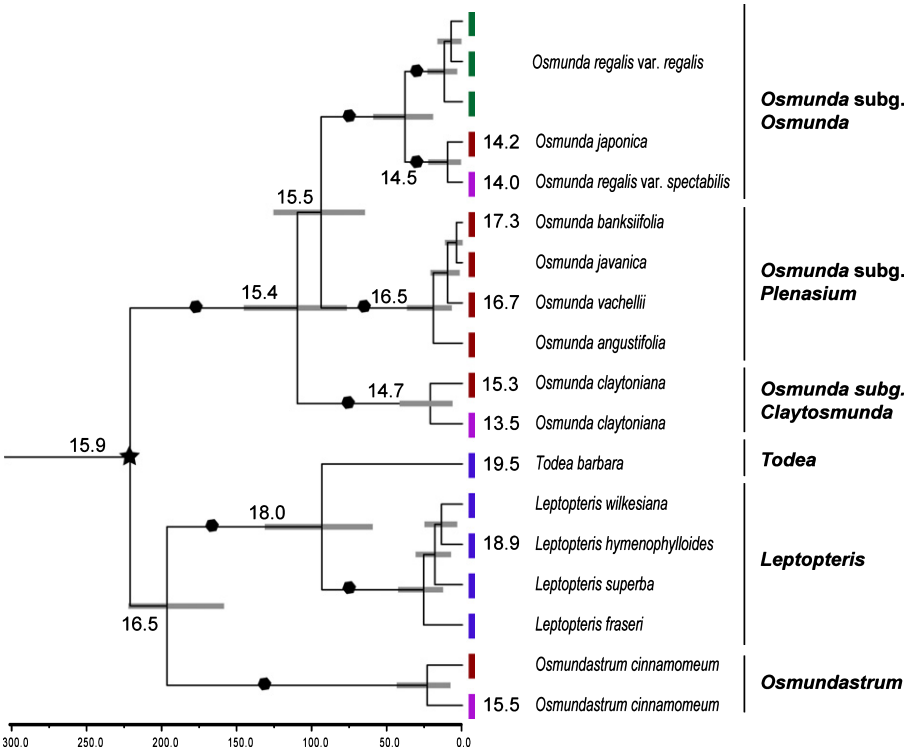


Fig. 1 Evolution of genome size in Osmundaceae through time as recovered by Bayesian diversification time estimates for extant royal ferns. The phylogenetic tree is based on *rbcL* sequences of all extant species used in previous studies (Yatabe *et al.*, 1999; Metzgar *et al.*, 2008; Tsutsumi *et al.*, 2011), including more than one sample of *Osmunda claytoniana*, *Osmunda regalis*, and *Osmundastrum cinnamomeum* to cover the known spatial differentiation of these species. We also included three samples of *Osmunda regalis* subsp. *regalis* representing occurrences in Africa, Europe, and India. Estimated 1C-values in picograms are given between terminal branches and species name, whereas reconstructed ancestral 1C-values are given above or below selected branches. The crown group (marked by a star) was calibrated with a lognormal distribution and an offset of 220 million yr based on Triassic *Osmundastrum*-like fossils (Phipps *et al.*, 1998). The time scale is given in millions of years before present. Confidence intervals are shown as grey horizontal bars whereas black hexagons mark branches with a posterior value of $P = 1.00$. Vertical lines at terminal branches indicate the geographical distribution: blue, southern hemisphere (mainly Australasia); green, Europe, Africa, India; purple, North America (and South America in the case of *Osmunda regalis*); red, Southeast Asia. Generic and subgeneric classifications are annotated at the right-hand side of the figure.

for the *rbcL* region of the plastid genome, if the chronogram is calibrated using existing fossils. This contrasts with an average substitution rate of 5.0×10^{-4} substitutions per site per million years estimated for the majority of ferns and land plants in general (Villarreal & Renner, 2014; H. Schneider unpublished). Even older ages and thus lower substitution rates are estimated if the

oldest fossils assigned to *Osmundastrum* and *Todea* (Wang *et al.*, 2013; Bomfleur *et al.*, 2014) are considered to belong to the crown group of these two extant genera rather than assigning the oldest royal fern fossil as an age constraint of the crown group of extant Osmundaceae (data not shown). Given the often arborescent sporophytes and the extended growing season of the gametophytes (Klekowski, 1973), it seems likely that royal ferns have long generation times. If so then these may contribute to the low substitution rates observed (as also suggested for tree ferns which have long generation times, Korall *et al.*, 2010; Zhong *et al.*, 2014), and perhaps also to the relatively limited genome size diversity (Beaulieu *et al.*, 2010).

Diversification rate

The earlier data are also complemented by the finding that net diversification rates of royal ferns are 2.44–4.43 times lower than the average rate reported for ferns as a whole. Indeed Osmundaceae have the lowest reported diversification rate of any leptosporangiate fern lineage (Schuettpelz & Pryer, 2009). However, this low rate is not associated with a complete absence of recent speciation as some clades are seen to have diverged since the late Oligocene (e.g. *Leptopteris*, *Osmunda* subg. *Plenasium* in Fig. 1). It is notable that the co-occurrence of relatively large genomes in royal ferns with a low diversification rate is consistent with some theoretical predictions (Kraaijeveld, 2010).

Overall, while these observations do suggest that the genomes of royal ferns are less dynamic than other fern lineages, they nevertheless challenge the hypothesis that their genomes are frozen in time. Instead of stasis, we find evidence for some variation suggesting rather slow evolution compared with other fern lineages. This is also reflected in various morphological features that are actually less static if all extant species of royal ferns are considered. For example, the lineage shows considerable variation in leaf morphology as illustrated by an absence of differentiation between sporangium-bearing and photosynthetically active parts in the leaves of *Leptopteris* and *Todea,* separation of these two functions within the leaf in *Osmunda* and between leaves in *Osmundastrum*. Once again, the explanation may lie in the long generation times of royal ferns, which have contributed to the limited speed of morphological evolution.

Despite our disagreement with the overall interpretation of Bomfleur *et al.* (2014), we agree with the unique opportunity provided by these ferns for studies aiming to understand how changes in genome dynamics have contributed to the evolution of plants and, in turn, the importance of genomic changes for the evolution of lineages. By analysing genomic structure in species of Osmundaceae, this will allow the exploration of genomic changes in a group of plants with obvious limitations in their evolutionary capacity. To interpret these findings, it is crucial to understand the unique phylogenetic position of royal ferns as they comprise the sister lineage of the most species-rich lineage of ferns (Pryer *et al.*, 2004). The royal ferns contain a unique combination of character states including those found only in these ferns (e.g. their stele) as well as characters that are intermediate between the plesiomorphic eusporangiate state and the apomorphic leptosporangiate state (Schneider *et al.*, 2009). It therefore seems likely that the genomes of royal ferns will display some characteristics shared with all ferns as well as those that are unique to these ferns. The recent proposal (Sessa *et al.*, 2014) to sequence two fern genomes is certainly welcomed, but it is noted that the genera selected (*Ceratopteris* and *Azolla*) occupy derived phylogenetic positions, well within the species-rich leptosporangiate clade. The addition of a member of Osmundaceae is strongly recommended, providing both an outgroup for comparative analyses with the *Ceratopteris* and *Azolla* data as well as the potential to gain fundamental insights into the genomics underpinning a lineage with limited evolutionary capacity.

Harald Schneider[1]*, Hongmei Liu[2], James Clark[1], Oriane Hidalgo[3], Jaume Pellicer[3], Shouzhou Zhang[2], Laura J. Kelly[4], Michael F. Fay[3] and Ilia J. Leitch[3]

[1]Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK;
[2]Key Laboratory of Southern Subtropical Plant Diversity, Fairylake Botanical Garden, Shenzhen & The Chinese Academy of Sciences, Shenzhen 518004, P. R. China;
[3]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW8 3DS, UK;
[4]School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK
(*Author for correspondence: tel +44 (0)2079426058; email h.schneider@nhm.ac.uk)

References

Arumuganathan K, Earle ED. 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9**: 208–218.

Bainard J, Henry T, Bainard L, Newmaster S. 2011. DNA content variation in monilophytes and lycophytes: large genomes that are not endopolyploid. *Chromosome Research* **19**: 763–775.

Beaulieu JM, Smith S, Leitch IJ. 2010. On the tempo of genome size evolution in angiosperms. *Journal of Botany* **2010**: 989152.

Bomfleur B, McLoughlin S, Vajda V. 2014. Fossilized nuclei and chromosomes reveal 180 million years of genomic stasis in Royal Ferns. *Science* **343**: 1376–1377.

Cheng RIJ, Grant WF. 1973. Species relationships in the *Lotus corniculatus* group as determined by karyotype and cytophotometric analyses. *Canadian Journal of Genetics and Cytology* **15**: 101–115.

Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, Obermayer R. 1998. Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Annals of Botany* **82** (Suppl. A): 17–26.

Doležel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* **85**: 625–631.

Ebihara A, Ishikawa H, Matsumoto S, Lin S-J, Iwatsuki K, Takamiya M, Watano Y, Ito M. 2005. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. *American Journal of Botany* **92**: 1535–1547.

Klekowski EJ. 1973. Sexual and subsexual systems in the homosporous ferns: a new hypothesis. *American Journal of Botany* **60**: 535–544.

Korall P, Schuettpelz E, Pryer KM. 2010. Abrupt deceleration of molecular evolution linked to the origin of arborescence in ferns. *Evolution* **64**: 2786–2792.

Kraaijeveld K. 2010. Genome size and species diversification. *Evolutionary Biology* **37**: 227–233.

Leitch AR, Leitch IJ. 2012. Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist* **194**: 629–646.

Leitch IJ, Leitch AR. 2013. Genome size diversity and evolution in land plants. In: Leitch IJ, Greilhuber J, Doležel J, Wendel JF, eds. *Plant genome diversity, vol 2, physical structure, behaviour and evolution of plant genomes.* Vienna, Austria: Springer-Verlag, 307–322.

Loureiro J, Rodriguez E, Doležel J, Santos C. 2007. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany* **100**: 875–888.

Manton I. 1950. *Problems of cytology and evolution in Pteridophyta.* Cambridge, UK: Cambridge University Press.

Metzgar JS, Skog JE, Zimmer EA, Pryer KM. 2008. The paraphyly of *Osmunda* is confirmed by phylogenetic analyses of seven plastid loci. *Systematic Botany* **33**: 31–36.

Obermayer R, Leitch IJ, Hanson L, Bennett MD. 2002. Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Annals of Botany* **90**: 209–217.

Pellicer J, Kelly LJ, Leitch IJ, Zomlefer WB, Fay MF. 2014. A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *New Phytologist* **201**: 1484–1497.

Phipps C, Taylor T, Taylor E, Cúneo R, Boucher L, Yao X. 1998. *Osmunda* (Osmundaceae) from the Triassic of Antarctica: an example of evolutionary stasis. *American Journal of Botany* **85**: 888.

Pryer KM, Schuettpelz E, Wolf PG, Schneider H, Smith AR, Cranfill R. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *American Journal of Botany* **91**: 1582–1598.

Schneider H, Smith AR, Pryer KM. 2009. Is morphology really at odds with molecules in estimating fern phylogeny? *Systematic Botany* **34**: 455–475.

Schuettpelz E, Pryer KM. 2009. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. *Proceedings of the National Academy of Sciences, USA* **106**: 11200–11205.

Senerchia N, Wicker T, Felber F, Parisod C. 2013. Evolutionary dynamics of retrotransposons assessed by high-throughput sequencing in wild relatives of wheat. *Genome Biology and Evolution* **5**: 1010–1020.

Serbet R, Rothwell GW. 1999. *Osmunda cinnamomea* (Osmundaceae) in the Upper Cretaceous of Western North America: additional evidence for exceptional species longevity among Filicalean ferns. *International Journal of Plant Sciences* **160**: 425–433.

Sessa E, Banks J, Barker M, Der J, Duffy A, Graham S, Hasebe M, Langdale J, Li F-W, Marchant D *et al.* 2014. Between two fern genomes. *GigaScience* **3**: 15.

Tsutsumi C, Matsumoto S, Yatabe-Kakugawa Y, Hirayama Y, Kato M. 2011. A new allotetraploid species of *Osmunda* (Osmundaceae). *Systematic Botany* **36**: 836–844.

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM *et al.* 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology* **30**: 83–89.

Villarreal JC, Renner SS. 2014. A review of molecular-clock calibrations and substitution rates in liverworts, mosses, and hornworts, and a timeframe for a taxonomically cleaned-up genus *Nothoceros*. *Molecular Phylogenetics and Evolution* **78**: 25–35.

Wang S-J, Hilton J, He X-Y, Seyfullah LJ, Shao L. 2013. The anatomically preserved stem *Zhongmingella* gen. nov. from the Upper Permian of China: evaluating the early evolution and phylogeny of the Osmundales. *Journal of Systematic Palaeontology* **12**: 1–22.

Yatabe Y, Nishida H, Murakami N. 1999. Phylogeny of Osmundaceae inferred from *rbcL* nucleotide sequences and comparison to the fossil evidences. *Journal of Plant Research* **112**: 397–404.

Zhang S-Z, He Z-C, Fan C-R, Yan B. 2008. A cytogenetic study of five species in the genus *Osmunda*. *Journal of Systematics and Evolution* **46**: 490–498.

Zhong BJ, Fong R, Collins LJ, McLenachan PA, Penny D. 2014. Two new fern chloroplasts and decelerated evolution linked to the long generation time in tree ferns. *Genome Biology and Evolution* **6**: 1166–1173.

Key words: diversification, evolutionary capacity, evolutionary stasis, fossil record, genome evolution, genome size, Osmundaceae.

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Puttick MN, Clark J, Donoghue PCJ. 2015 Size is not everything: rates of genome size evolution, not *C*-value, correlate with speciation in angiosperms. *Proc. R. Soc. B* **282**: 20152289. http://dx.doi.org/10.1098/rspb.2015.2289

Received: 21 September 2015
Accepted: 3 November 2015

Subject Areas:
evolution, plant science, genetics

Keywords:
angiosperms, genome size, evolvability, polyploidy, genome duplication

Author for correspondence:
Mark N. Puttick
e-mail: mark.puttick@bristol.ac.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2015.2289 or via http://rspb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Size is not everything: rates of genome size evolution, not *C*-value, correlate with speciation in angiosperms

Mark N. Puttick, James Clark and Philip C. J. Donoghue

School of Biological Sciences, University of Bristol, Bristol Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK

Angiosperms represent one of the key examples of evolutionary success, and their diversity dwarfs other land plants; this success has been linked, in part, to genome size and phenomena such as whole genome duplication events. However, while angiosperms exhibit a remarkable breadth of genome size, evidence linking overall genome size to diversity is equivocal, at best. Here, we show that the rates of speciation and genome size evolution are tightly correlated across land plants, and angiosperms show the highest rates for both, whereas very slow rates are seen in their comparatively species-poor sister group, the gymnosperms. No evidence is found linking overall genome size and rates of speciation. Within angiosperms, both the monocots and eudicots show the highest rates of speciation and genome size evolution, and these data suggest a potential explanation for the megadiversity of angiosperms. It is difficult to associate high rates of diversification with different types of polyploidy, but it is likely that high rates of evolution correlate with a smaller genome size after genome duplications. The diversity of angiosperms may, in part, be due to an ability to increase evolvability by benefiting from whole genome duplications, transposable elements and general genome plasticity.

1. Introduction

Evolutionary biology has long sought to explain the uneven diversity across the branches of the tree of life. The land plants (Embryophyta) are a focal example, with approximately 320 000 species known, 268 600 are angiosperms [1]; indeed, the immediate sister lineage of angiosperms can muster only approximately 1050 species [1]. Many factors have been used to explain this imbalance, such as environmental opportunity [2] and key adaptations [3,4], whereas recent attention has been focused on genome size [5–7].

Across the tree of life, genome size has been linked causally to increased diversification. Traditionally, larger genomes have been linked to greater rates of speciation, but there is also evidence of smaller genomes promoting diversification, including in plants [8–10]. Furthermore, many factors relating to genome size are related to higher diversification in plants: whole genome duplication [5,11–18], transposable elements [7] and selective pressures can cause differences in genome size and diversification [10]. Theory and some experimental evidence suggests a role for genome size in variations of diversification rates, but much attention has so far has concentrated upon the size of genomes, yielding equivocal results [10].

Angiosperms are exceptional in their approximately 2000-fold variation in genome size, which has been linked to their successful diversification [5,19,20]. This contrasts strongly with the narrow variance in the larger genomes of gymnosperms [5,12,21,22]. Many factors related to evolvability are expected to alter genome size, but not unidirectionally towards a larger or smaller size [23]. Therefore, rates of size change, not absolute size, of genomes, are

likely to be an important factor in explaining the differing rates of diversification across land plants.

High rates of trait evolution are associated with increased diversification potential across the tree of life [24,25]. High rates of genome size evolution promoting higher diversification in angiosperms are compatible with this hypothesis. Two main theories could explain a positive relationship between the two: punctuated evolution, in which the majority of phenotypic change occurs at speciation [26,27], especially in plants where there is a high incidence of polyploidy [28], or some form of 'evolvability', in which the capacity to change phenotype allows for higher rates of speciation [24,25]. However, differentiating punctuational models from evolvability models can be difficult [29], and it is likely the two are not mutually exclusive.

Genome size evolution can be modelled as a trait on a phylogenetic tree, and this allows for testing of the correlation between the rates of diversification and genome size evolution [30,31]. Here, we test this relationship across land plants using a large database of genome sizes, and predict a positive correlation between high rates of genome size evolution and speciation across the phylogeny, particularly in the angiosperms, but expect no relationship with genome size and speciation. We find this relationship to be true, with particularly high levels of size evolution in the eudicots and monocots, particularly the grasses (Poaceae). The ability to rapidly change genome size may have increased the evolvability of angiosperms, and allowed them to diversify spectacularly.

2. Methods

The most comprehensive, dated phylogeny of land plants [32] was used to model genome size evolution. When genome size data were considered, the phylogeny was pruned down to 3351 species of land plants.

We obtained genome sizes (1C, picograms) from the Kew C-value database [19]. Although we term 1C as 'genome size' here, we recognize the true definition is of 2C divided by the level of ploidy [30,33].

(a) Rates of speciation and genome size evolution

Bayesian analysis of macroevolutionary mixtures (BAMMs) was used to analyse genome size evolution and rates of speciation separately on the phylogeny [25,34]. BAMM allows for multiple rate shift configurations to be modelled on phylogenies, thus it is not dependent upon a single shift configuration. Rate shifts are modelled via a compound Poisson process [34], and so no priors are required on the location of rate shifts. Diversification is modelled using parameters to represent speciation and extinction, and trait evolution is modelled as a Brownian motion process [25,34].

Priors for the reversible-jump mcmc model in BAMM were estimated using BAMMtools [35] in the software package R [36]. BAMM was run for 400 million generations for the phenotypic data, and 40 million for the analyses of speciation. Convergence was judged upon parameters exceeding 200 estimated sample size; this was more than 1000 for most parameters in the phenotypic data and analyses of speciation.

To incorporate non-random incomplete sampling, we followed established BAMM protocols. We assigned each species to a monophyletic family and calculated the proportion of species present in each family, as well as the overall proportion of land plant species. We obtained information about the number of valid species, as well as total plant species, from the plant list [37].

(b) Correlation between rates of genome size evolution and speciation

Correlation between the rates of genome size evolution and speciation within 276 embryophyte families [25], and rates were estimated for higher-level clades. The second was to study correlations between the rate of phenotypic evolution and family diversity, in terms of species richness [38,39]. We also tested whether size was correlated with speciation rates across the tree using traitDependent BAMM, which is a method that computes correlation coefficients between the trait and random posterior speciation rates from BAMM samples.

Phylogenetic generalized least-squares (GLS) models were used to account for the effects of phylogeny in the regression of speciation rates on rates of genome size evolution [25,39–41]. PGLS models were based on code from the CAPER package in R [42]. PGLS quantifies and incorporates similarity between species owing to the shared phylogenetic history by estimating Pagel's λ [40,43]—this similarity is then incorporated into the error term of the regression model [44].

As we tested the correlations of two rates, both could be positively correlated with time [25]. Therefore, we also tested for evidence of this relationship by looking at the influence of time by examining the rates between sister-clades only which, by definition, are of equal age [25].

(c) Direction of change

We used StableTraits [45] to estimate ancestral sizes of genomes throughout the phylogeny. StableTraits samples rates from a heavy-tailed [45,46], rather than a normal distribution, as in Brownian motion [47]. This allows for rate changes to be estimated parametrically on the tree, such that individual branch rates and ancestral node estimates can be calculated for the entire tree. StableTraits was run for 80 million generations, sampling at every 1000 generations, and across two independent chains.

3. Results

(a) Rates of speciation and genome size evolution

Speciation and genome size evolution show considerable variation throughout the phylogeny. In the model of genome rate evolution, the mean log-likelihood of the posterior was 3583.77 (3426.84–3740.07, 2.5 and 97.5 percentiles, respectively) and the mean number of shifts was 62 (56–69, 2.5 and 97.5 percentiles, respectively). Similar results were found for rates of speciation: the mean number of shifts was 48 (39–58, 0.025 and 0.975 quantiles, respectively), and the mean log-likelihood of the posterior was −11 534.65 (−11 674.6 to −11 448.8, 02.5 and 97.5 percentiles, respectively). Although it was not possible to calculate Bayes factors—the prior was zero for many of the shifts—there is a clear difference between the prior and posterior for the number of shifts (see electronic supplementary material, figures S1 and S2).

Angiosperms show the highest rates of genome size evolution and speciation (table 1 and figure 1). Mean clade rates in the angiosperms for speciation (0.55) and genome size evolution (0.009) were higher compared with the speciation rate (0.04) and genome size evolution rates (0.001) in non-angiosperms. Within angiosperms, very high rates of genome size evolution are found within monocots (figure 1), particularly Poaceae (0.16), which also exhibits the highest rate of speciation (4.53). The lowest rates of speciation (0.03) and genome evolution (0.03) are found in gymnosperms. The

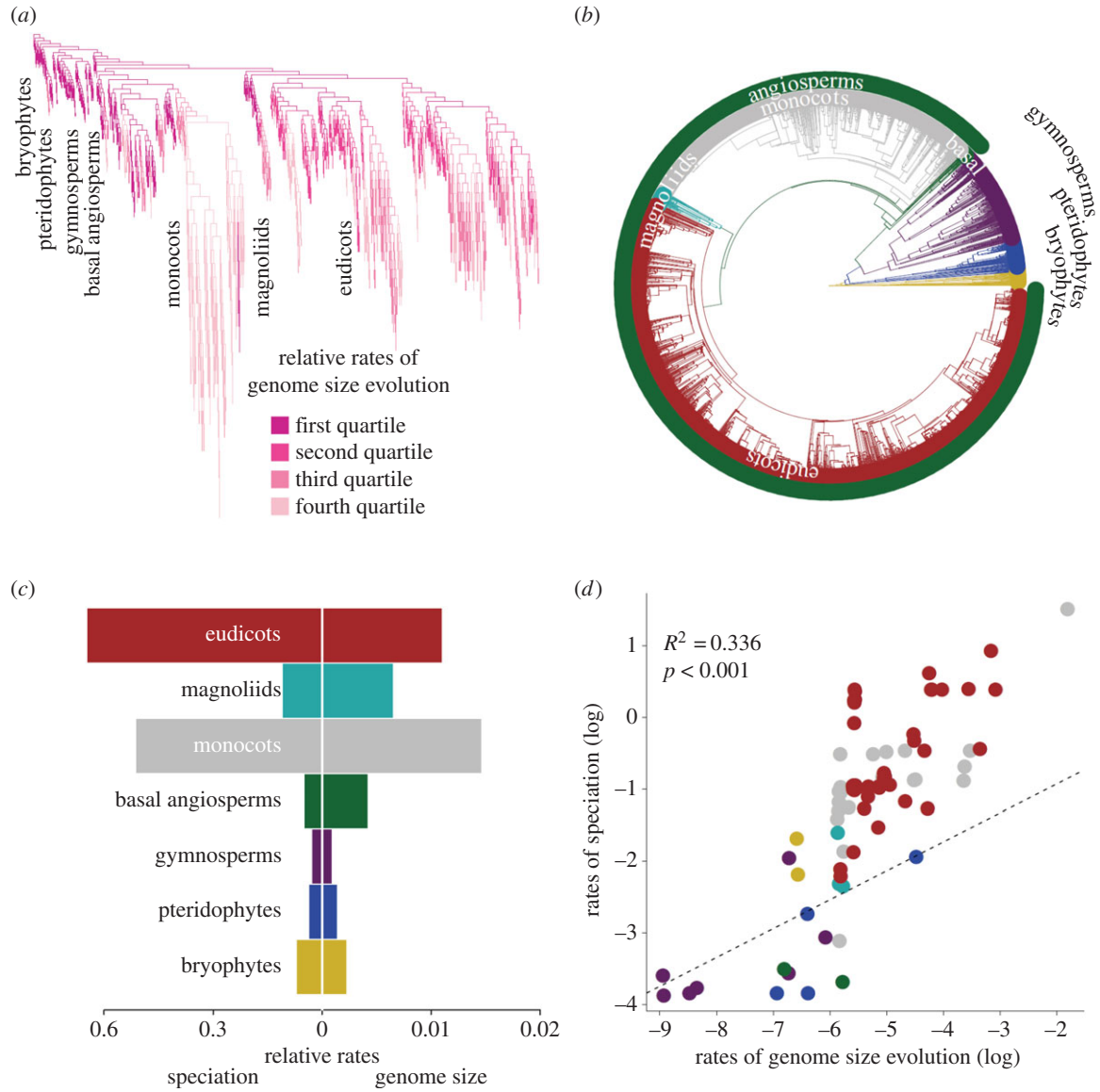


Figure 1. Rates of speciation and genome evolution are correlated in plants. The highest rates of speciation (branches scaled to rate) are associated with the highest genome rates (coloured branches) (*a*). Clades shown in the phylogeny (*b*) show correlation between rates of genome size evolution and speciation (*c*), and there is a significant relationship in a phylogenetically corrected correlation between the two rates for families (*d*).

Table 1. Rates of speciation and genome size evolution for clades in the phylogeny.

	speciation rate	genome rate
angiosperms	0.55 (0.48, 0.62)	0.009 (0.008, 0.01)
non-angiosperms	0.04 (0.03, 0.07)	0.001 (0.001, 0.001)
bryophytes	0.07 (0.04, 0.18)	0.002 (0.0009, 0.004)
pteridophytes	0.04 (0.03, 0.06)	0.001 (0.0006, 0.002)
gymnosperms	0.03 (0.02, 0.04)	0.0007 (0.0003, 0.002)
basal angiosperms	0.05 (0.02, 0.27)	0.003 (0.002, 0.005)
magnoliids	0.11 (0.05, 0.33)	0.005 (0.002, 0.01)
monocots	0.51 (0.42, 0.65)	0.011 (0.009, 0.01)
eudicots	0.65 (0.52, 0.72)	0.008 (0.007, 0.009)

families Pinaceae (0.0001) and Araucariaceae (0.02) have the lowest speciation and genome rates, respectively (see electronic supplementary material, figure S4).

(b) Positive correlation between rates of genome size evolution and speciation

At the family level, there is a significant relationship between rates of genome size evolution and speciation across the tree (figure 1). The PGLS model, which tests for the significance of the relationship at the family level (figure 1*b*,*d*), indicates a strong relationship between genome size evolution and speciation rates ($p < 0.001$, 90 d.f., $R^2 = 0.383$). This is also significant within just angiosperms ($p < 0.001$, 76 d.f., $R^2 = 0.359$) (table 2). These results are also significant when using contrasts.

As an analogous test, the relationship between tip diversity of families (*n* species) and rates of genome size evolution was performed. This was very significant for the entire tree ($p < 0.001$, 90 d.f., $R^2 = 0.357$) and within just angiosperms ($p < 0.001$, 76 d.f., $R^2 = 0.219$; table 2 and electronic supplementary material, figure S3*a*,*b*).

Independent contrast also gave similar results to PGLS with a significant relationship between the genome size and speciation rates ($p < 0.001$, rho = 0.61). Time does not appear to be a confounding factor as contrasts between sister-species only was non-significant using the Spearman

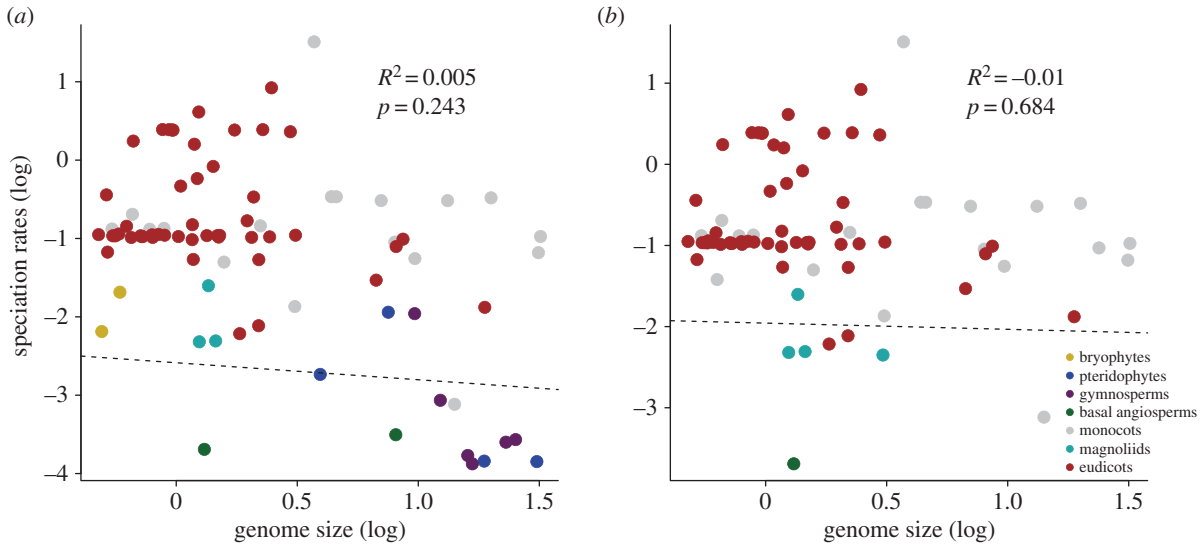


Figure 2. There is no significant relationship between overall genome size and rates of speciation for all land plants (*a*), and just angiosperms (*b*) when using a PGLS regression at the family level. Permutation tests also show a non-significant relationship between genome size and speciation for all plants.

Table 2. PGLS analyses show the positive relationship between genome size rates of evolution and speciation rates and family diversity for all plants and angiosperms only.

	d.f.	*p*-value	R^2	lambda (95% CIs)
all plants				
speciation rates	90	3.02×10^{-11}	0.3826	0.593 (n.a., 0.895)
family diversity	90	2.08×10^{-10}	0.3565	0 (0, 0.408)
angiosperms only				
speciation rates	76	1.62×10^{-8}	0.336	1 (0.874, n.a.)
family diversity	76	9.19×10^{-6}	0.2192	0 (0, 0.496)

rank test ($p = 0.054$). While this is used to test the confounding effect of time on analyses [25], it is likely that our negative result here is due to the small sample size ($n = 28$), and there is still a positive relationship (rho = 0.37). Furthermore, gymnosperms and angiosperms are the same age, by definition, and show no evidence of correlation in rates.

There is no evidence for high rates of speciation being linked to genome size (as opposed to *rates* of genome size evolution; figure 2). We find no significant correlation between overall speciation rates and genome size for the entire tree ($p = 0.243$, 83 d.f., $R^2 = 0.005$), or angiosperms ($p = 0.68$, 76 d.f., $R^2 = -0.01$). traitDependentBAMM also shows a non-significant correlation between genome size and speciation rates across the tree ($p = 0.56$).

We find little evidence for accelerations on branches leading to the major clades of angiosperms at sites associated with whole genome duplications. Rates on branches leading to angiosperms (0.003), monocots (0.002) and eudicots (0.003) all fall into the first quartile of rates throughout the phylogeny. Furthermore, there is little evidence to link purported whole genome size changes and accelerated rates of speciation or genome size evolution. We plotted the posited location of whole genome duplication events on the phylogenies displaying the best shift configurations of diversification and genome size evolution, respectively (minimum Bayes factor 5); these results indicate that only the core eudicots are associated with a shift in speciation and trait evolution rates (figure 3). Other whole genome duplication events are not associated with differences in speciation and trait evolution rates of evolution.

(c) Ancestral states and the direction of change

The reconstructed ancestral angiosperm genome size is 1.45 picograms (0.57–3.71 95% highest posterior density) which is smaller than the size estimated for the ancestral spermatophyte of 1.99 picograms (0.7105.49 95% highest posterior density; see electronic supplementary material, table S1 and figure 4). As expected, high rates of genome size evolution are associated with increases and decreases in *C*-value throughout the tree; there is no difference in the distribution of size changes in ancestor–descendant pairs between angiosperms and non-angiosperms ($p = 0.1531$, Wilcoxon rank-sum test). Therefore, it appears increased rates are associated with both increases and decreases in *C*-value throughout the phylogeny.

4. Discussion

While genome size has been traditionally linked to the success of angiosperms, here we find that it is the ability to alter genome size that exhibits the strongest correlation with diversity. This fits a hypothesis in which genome size in and of itself is not an important factor for diversification

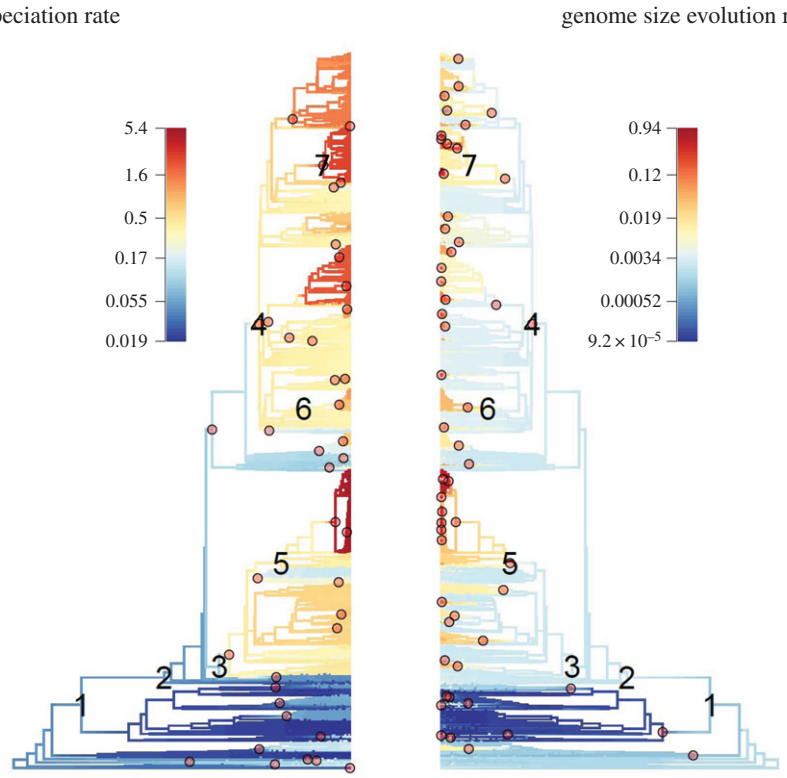


Figure 3. The position of shifts for rates of speciation and genome size evolution on the phylogeny compared with reported whole genome duplications in the Spermatophyta (1), Angiospermae (2), monocots (3), eudicots (4), Poaceae (5), Brassicaceae (6) and the Asteraceae (7). Only the core eudicots (4) show accelerated rates for speciation and genome size evolution.

as has been previously suggested [10], but it is the ability to cope with genome size changes that has allowed angiosperms to benefit from polyploidy and other genome rearrangements [5,8,9,12,48]. Changes in genome size are likely to have promoted diversification in angiosperms, especially compared with the species-poor gymnosperms [22].

As expected, the large variance in *C*-value for angiosperms [5,12,49] translates into a high rate of genome size evolution, and this correlates strongly with rates of speciation (figure 1). A frequent explanation for the huge diversity of angiosperms is the prevalence of whole genome duplication events [5,20]. However, directly linking *C*-value to polyploidy events can be difficult: *C*-value is not directly proportional to ploidy and often downsizes following duplication [50,51]. As we measure changes in *C*-value, these are very likely to be influenced by whole genome duplications as well as other factors linked to increased rates of diversification, such as tandem duplications, transposable elements ([7,47], but see [52]), life history [53] and deletions [8,51,54]). As a guide to 'genome size', *C*-value effectively captures large-scale patterns in genome size change throughout the phylogeny, but it is not attributable to one effect, such as whole genome duplications, alone. Overall, we support a model in which higher rates of genome size evolution that result from range of processes promote higher rates of speciation [7] (figure 1).

(a) Evolvability

High rates of genome size evolution correlate with high rates of speciation in angiosperms, and confirm previous predictions that genome size variability is linked to success in flowering plants [5]. These patterns could fit a punctuational model of evolution in which genome size changes occur at speciation [26], or a model of evolvability in which higher rates of genome change drives high rates of speciation [7,24,25]. Discriminating among punctuational and evolvability models is not trivial [29], and we cannot reject the possibility that they are linked, but this does not require one model being favoured at the expense of another. A large amount of change may be expected at speciation in a punctuational model [7,26–28,55]. A subset of this model posits that genome size changes, and by definition, speciation, are associated with cladogenesis—speciation results from polyploidy, but polyploidy does not promote diversification [12,28,49]. These models would imply small genome size is a consequence of, not a driving factor behind, diversification. However, we find no link between genome size and rates of speciation (figure 2), and we expect to find a small genome size in many species that have undergone recent, rapid radiations [5,56]. Therefore, there are many reasons to associate genome size change with higher rates of speciation in an evolvability model (figure 1): whole genome duplications [13,14], via general genome plasticity [5,12,48], lowering extinction risk by reducing genome size [8], the action of transposable elements [7] and retaining benefits of duplicated genes [48]. Thus, we cannot definitively differentiate between punctuational and evolvability models, but we suggest there is evidence to infer an evolvability model relating to higher rates of genome size evolution in plants (figure 1).

(b) Whole genome duplications

In the past, authors have argued that polyploidy and duplicated elements within genomes could lead to 'genetic

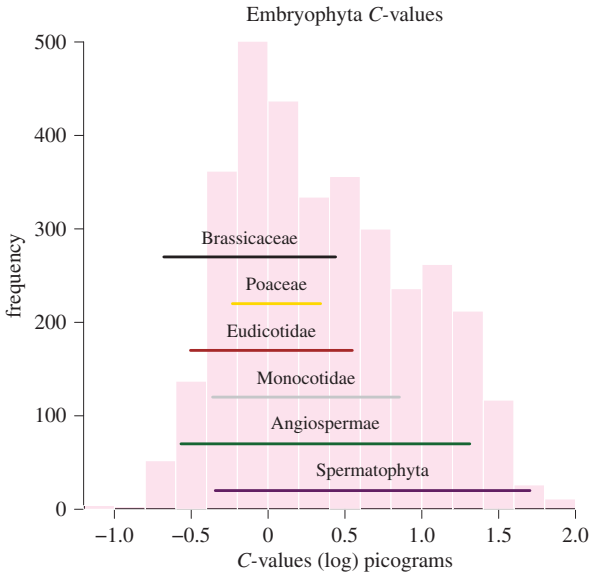


Figure 4. There is large uncertainty in ancestral reconstruction of genome size for nodes associated with whole genome duplication events. The histogram shows known *C*-values from extant land plants, and the coloured lines represent the range of uncertainty (95th% highest posterior density) for ancestral reconstruction of genome size in the common ancestors of crown clades.

obesity' [57], but despite multiple rounds of duplication we find no evidence for directional evolution in genome size. While it has become clear that increases and decreases in genome size are characteristic of angiosperms [5,30,51], we find no relationship between absolute genome size and rates of speciation in angiosperms or in embryophytes more generally (figure 2). Out of a number of proposed genome duplications [16,58–62], only core eudicots show a consistent shifts in rate for genome size evolution and diversification (as judged by Bayes factors; figure 3), and some clades associated with ancestral polyploidy show heightened rates of diversification (monocots, eudicots, Brassiceae, Asteraceae and Poaceae). Spermatophyta and Angiospermae do not show heightened speciation or genome size evolution rates. It can be seen that not all angiosperms have experienced a heightened rate of evolution (figure 1). This might evidence a model in which early-diverging lineages, including *Amborella*, did not undergo recent rounds of whole genome duplication and so do not exhibit higher rates of speciation [63], and demonstrates how nested diversifications may follow from whole genome duplications [20]. A relatively small ancestral angiosperm genome size has been suggested [64], but here the posterior density around our estimates for ancestral angiosperms is very large (figure 4). At present, it is possible to elucidate large-scale patterns in genome size evolution, but obtaining precise ancestral estimates for angiosperms may be difficult [65,66], but promise may come through working with fossils ([67], but see also [68]).

(c) Auto- and allopolyploidy

In this study, we do not differentiate between auto- and allopolyploidy, and the related subject of dosage-dependent and dosage-independent genes. Autopolyploidy is initially thought to maintain dosage balance via the retention of dosage-dependent genes, though over time it is thought that these may diverge in function or expression [23,69]. However, genomic rearrangements and heterosis effects are thought to be stronger in allopolyploids [69], and so it is likely to have had a large role in plant evolution, but current methods only tentatively identify a small number of differentiable auto- and allopolyploidy events ($n = 9$), and some of these are not phylogenetically positioned [69]. Thus, making statistical analysis of these events unfeasible at present, but incorporation of auto- and allopolyploidy events will improve future investigations.

5. Conclusion

Rates of genome size evolution are positively correlated with diversification rates in plants, a trend that is driven by largely by the positive relationship in angiosperms. No evidence supports a link between overall size and diversification. Overall, these results support a model in which rate of genome size evolution promotes the acquisition of novel traits, reproductive barriers and movement into new niches, which have aided the diversification of angiosperms.

Authors' contributions. M.N.P. devised the project, M.N.P. carried out the analyses and M.N.P., J.C. and P.C.J.D. wrote the manuscript.

Funding. We thank NERC for providing grant no. NE/K500823/1 to M.N.P.

Competing interests. We declare we have no competing interests.

Acknowledgements. We are grateful to members of the Bristol Palaeobiology Group for suggestions that improved the manuscript. Phylogenetic analyses were carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (http://www.bris.ac.uk/acrc/).

References

1. Chapman AD. 2009 *Numbers of living species in Australia and the world*, 2nd edn. Report for the Australian Biological Resources Study, Canberra, Australia. See https://www.environment.gov.au/system/files/pages/2ee3f4a1-f130-465b-9c7a-79373680a067/files/nlsaw-2nd-complete.pdf.
2. Hughes C, Eastwood R. 2006 Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl Acad. Sci. USA* **103**, 10 334–10 339. (doi:10.1073/pnas.0601928103)
3. Hodges SA, Arnold ML. 1995 Spurring plant diversification: are floral nectar spurs a key innovation? *Proc. R. Soc. Lond. B* **262**, 343–348. (doi:10.1098/rspb.1995.0215)
4. Hodges SA. 1997 Floral nectar spurs and diversification. *Int. J. Plant Sci.* **158**, S81–S88. (doi:10.1086/297508)
5. Leitch AR, Leitch IJ. 2008 Genomic plasticity and the diversity of polyploid plants. *Science* **320**, 481–483. (doi:10.1126/science.1153585)
6. Leitch IJ, Beaulieu JM, Chase MW, Leitch AR, Fay MF. 2010 Genome size dynamics and evolution in monocots. *J. Bot.* **2010**, 1–18. (doi:10.1155/2010/862516)
7. Oliver KR, McComb JA, Greene WK. 2013 Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* **5**, 1886–1901. (doi:10.1093/gbe/evt141)
8. Vinogradov AE. 2003 Selfish DNA is maladaptive: evidence from the plant Red List. *Trends Genet.* **19**, 609–614. (doi:10.1016/j.tig.2003.09.010)
9. Knight CA, Molinari NA, Petrov DA. 2005 The large genome constraint hypothesis: evolution, ecology

and phenotype. *Ann. Bot*. **95**, 177–190. (doi:10.1093/aob/mci011)

10. Kraaijeveld K. 2010 Genome size and species diversification. *Evol. Biol*. **37**, 227–233. (doi:10.1007/s11692-010-9093-4)
11. De Bodt S, Maere S, Van De Peer Y. 2005 Genome duplication and the origin of angiosperms. *Trends Ecol. Evol*. **20**, 591–597. (doi:10.1016/j.tree.2005.07.008)
12. Leitch AR, Leitch IJ. 2012 Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytol*. **194**, 629–646. (doi:10.1111/j.1469-8137.2012.04105.x)
13. Fawcett JA, Maere S, Van de Peer Y. 2009 Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl Acad. Sci. USA* **106**, 5737–5742. (doi:10.1073/pnas.0900906106)
14. Van de Peer Y, Maere S, Meyer A. 2009 The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet*. **10**, 725–732. (doi:10.1038/nrg2600)
15. Murat F, Van De Peer Y, Salse J. 2012 Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biol. Evol*. **4**, 917–928. (doi:10.1093/gbe/evs066)
16. Kagale S *et al.* 2014 Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* **26**,2777–2791. (doi:10.1105/tpc.114.126391)
17. Vanneste K, Baele G, Maere S, Van De Peer Y. 2014 Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res*. **24**, 1334–1347. (doi:10.1101/gr.168997.113)
18. Edger PP *et al.* 2015 The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl Acad. Sci. USA* **112**, 8362–8366. (doi:10.1073/pnas.1503926112)
19. Bennett MD, Leitch IJ, 2012 Plant DNA C-values database (release 6.0, December 2012). See http://www.kew.org/cvalues/.
20. Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchcliff CE, Brown JW, Sessa EB, Harmon LJ. 2015 Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol*. **207**, 454–467. (doi:10.1111/nph.13491)
21. Leitch IJ, Soltis DE, Soltis PS, Bennett MD. 2005 Evolution of DNA amounts across land plants (Embryophyta). *Ann. Bot*. **95**, 207–217. (doi:10.1093/aob/mci014)
22. Nystedt B *et al.* 2013 The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584. (doi:10.1038/nature12211)
23. Conant GC, Birchler JA, Pires JC. 2014 Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol*. **19**, 91–98. (doi:10.1016/j.pbi.2014.05.008)
24. Pigliucci M. 2008 Is evolvability evolvable? *Nat. Rev. Genet*. **9**, 75–82. (doi:10.1038/nrg2278)
25. Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME. 2013 Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun*. **4**, 1–8. (doi:10.1038/ncomms2958)
26. Eldredge N, Gould SJ. 1972 Punctuated equilibria: an alternative to phyletic gradualism. In *Models in paleobiology* (ed. TJM Schopf), pp. 82–115. San Francisco, CA: Freeman and Cooper. (doi:10.1037/h0022328)
27. Pennell MW, Harmon LJ, Uyeda JC. 2014 Is there room for punctuated equilibrium in macroevolution? *Trends Ecol. Evol*. **29**, 23–32. (doi:10.1016/j.tree.2013.07.004)
28. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009 The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA* **106**, 13 875–13 879. (doi:10.1073/pnas.0811575106)
29. Rabosky DL. 2012 Positive correlation between diversification rates and phenotypic evolvability can mimic punctuated equilibrium on molecular phylogenies. *Evolution* (*NY*) **66**, 2622–2627. (doi:10.1111/j.1558-5646.2012.01631.x)
30. Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003 Evolution of genome size in the angiosperms. *Am. J. Bot*. **90**, 1596–1603. (doi:10.3732/ajb.90.11.1596)
31. Burleigh JG, Barbazuk WB, Davis JM, Morse AM, Soltis PS. 2012 Exploring diversification and genome size evolution in extant gymnosperms through phylogenetic synthesis. *J. Bot*. **2012**, 1–6. (doi:10.1155/2012/292857)
32. Zanne AE *et al.* 2014 Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92. (doi:10.1038/nature12872)
33. Bennett MD, Leitch IJ, Hanson L. 1998 DNA amounts in two samples of angiosperm weeds. *Ann. Bot*. **82**, 121–134. (doi:10.1006/anbo.1998.0785)
34. Rabosky DL. 2014 Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543. (doi:10.1371/journal.pone.0089543)
35. Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Huang H, Larson JG. 2014 BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol*. **5**, 701–707. (doi:10.1111/2041-210X.12199)
36. R Development Core Team. 2012 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. R Found. Stat. Comput.
37. The Plant List. 2010 version 1. Published on the Internet 2010; See http://www.theplantlist.org/.
38. Rabosky DL, Adams DC. 2012 Rates of morphological evolution are correlated with species richness in salamanders. *Evolution* (*NY*) **66**, 1807–1818. (doi:10.1111/j.1558-5646.2011.01557.x)
39. Zelditch ML, Li J, Tran LAP, Swiderski DL. 2015 Relationships of diversity, disparity, and their evolutionary rates in squirrels (Sciuridae). *Evolution* (*NY*) **69**, 1284–1300. (doi:10.1111/evo.12642)
40. Pagel M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884. (doi:10.1038/44766)
41. Freckleton RP, Harvey PH, Pagel M. 2002 Phylogenetic analysis and comparative data. *Am. Nat*. **160**, 712–726. (doi:10.1086/343873)
42. Orme CDL. 2012 The caper package: comparative analysis of phylogenetics and evolution in R. See http://cran.r-project.org/package=caper.
43. Pagel M. 1997 Inferring evolutionary processes from phylogenies. *Zool. Scr*. **26**, 331–348. (doi:10.1111/j.1463-6409.1997.tb00423.x)
44. Revell LJ. 2010 Phylogenetic signal and linear regression on species data. *Methods Ecol. Evol*. **1**, 319–329. (doi:10.1111/j.2041-210X.2010.00044.x)
45. Elliot M, Mooers A. 2014 Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *BMC Evol. Biol*. **14**, 226. (doi:10.1186/s12862-014-0226-8)
46. Landis MJ, Schraiber JG, Liang M. 2013 Phylogenetic analysis using Levy processes: finding jumps in the evolution of continuous traits. *Syst. Biol*. **62**, 193–204. (doi:10.1093/sysbio/sys086)
47. Felsenstein J. 1985 Phylogenies and the comparative method. *Am. Nat*. **125**, 1–15. (doi:10.1086/284325)
48. Fedoroff NV. 2012 Transposable elements, epigenetics, and genome evolution. *Science*. **338**, 758–767. (doi:10.1126/science.338.6108.758)
49. Kejnovsky E, Leitch IJ, Leitch AR. 2009 Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends Ecol. Evol*. **24**, 572–582. (doi:10.1016/j.tree.2009.04.010)
50. Leitch IJ, Bennett MD. 2004 Genome downsize in polyploids plants. *Biol. J. Linn. Soc*. **82**, 651–663. (doi:10.1111/j.1095-8312.2004.00349.x)
51. Hufton AL, Panopoulou G. 2009 Polyploidy and genome restructuring: a variety of outcomes. *Curr. Opin. Genet. Dev*. **19**, 600–606. (doi:10.1016/j.gde.2009.10.005)
52. Fehér T *et al.* 2012 Competition between transposable elements and mutator genes in bacteria. *Mol. Biol. Evol*. **29**, 3153–3159. (doi:10.1093/molbev/mss122)
53. Beaulieu JM, Smith SA, Leitch IJ. 2010 On the tempo of genome size evolution in angiosperms. *J. Bot*. **2010**, 1–8. (doi:10.1155/2010/989152)
54. Otto SP. 2007 The evolutionary consequences of polyploidy. *Cell* **131**, 452–462. (doi:10.1016/j.cell.2007.10.022)
55. Mallet J. 2007 Hybrid speciation. *Nature* **446**, 279–283. (doi:10.1038/nature05706)
56. Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA. 2013 The fate of duplicated genes in a polyploid plant genome. *Plant J*. **73**, 143–153. (doi:10.1111/tpj.12026)
57. Bennetzen J, Kellogg E. 1997 Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**, 1509–1514. (doi:10.1105/tpc.9.9.1509)

58. Jiao Y *et al.* 2011 Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100. (doi:10.1038/nature09916)
59. Tang H, Bowers JE, Wang X, Paterson AH. 2010 Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl Acad. Sci. USA* **107**, 472–477. (doi:10.1073/pnas.0908007107)
60. Jiao Y, Li J, Tang H, Paterson AH. 2014 Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* **26**, 1–12. (doi:10.1105/tpc.114.127597)
61. Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH. 2008 Multiple paleopolyploidizations during the evolution of the composite reveal parallel patterns of duplicate gene retention after millions of years. *Mol. Biol. Evol*. **25**, 2445–2455. (doi:10.1093/molbev/msn187)
62. Soltis DE *et al.* 2009 Polyploidy and angiosperm diversification. *Am. J. Bot*. **96**, 336–348. (doi:10.3732/ajb.0800079)
63. Albert VA *et al.* 2013 The *Amborella* genome and the evolution of flowering plants. *Science*. **342**, 1241089. (doi:10.1126/science.1241089)
64. Leitch I, Leitch I. 1998 Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann. Bot*. **82**, 85–94. (doi:10.1006/anbo.1998.0783)
65. Oakley TH, Cunningham CW. 2000 Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution* **54**, 397–405. (doi:10.1111/j.0014-3820.2000.tb00042.x)
66. Webster AJ, Purvis A. 2002 Testing the accuracy of methods for reconstructing ancestral states of continuous characters. *Proc. R. Soc. Lond. B* **269**, 143–149. (doi:10.1098/rspb.2001.1873)
67. Bomfleur B, Mcloughlin S, Vajda V. 2014 Genomic stasis in royal ferns. *Science* **343**, 1376–1377.
68. Schneider H, Liu H, Clark J, Hidalgo O, Pellicer J, Zhang S, Kelly LJ, Fay MF, Leitch IJ. 2015 Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in the royal ferns. *New Phytol*. **207**, 10–13. (doi:10.1111/nph.13330)
69. Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2014 Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol*. **31**, 448–454. (doi:10.1093/molbev/mst230)

New Phytologist

Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny

James Clark[1,2]*, Oriane Hidalgo[3]*, Jaume Pellicer[3], Hongmei Liu[4], Jeannine Marquardt[1], Yannis Robert[5], Maarten Christenhusz[3,6], Shouzhou Zhang[4], Mary Gibby[7], Ilia J. Leitch[3] and Harald Schneider[1,8]

[1]Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK; [2]School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK; [3]Royal Botanic Gardens, Kew, Richmond, Surrey, TW8 3DS, UK; [4]Shenzhen Key Laboratory of Southern Subtropical Plant Diversity, Fairylake Botanical Garden, Shenzhen & The Chinese Academy of Sciences, Shenzhen 518004, China; [5]18, Rue des Capucines, F-97431, La Plaine des Palmistes, La Réunion, France; [6]Plant Gateway, 5 Talbot Street, Hertford, Hertfordshire, SG13 7BX, UK; [7]Department of Science, Royal Botanic Garden Edinburgh, Edinburgh, EH3 5LR, UK; [8]School of Life Sciences, Sun Yatsen University, Guangzhou 510275, Guangdong, China

Author for correspondence:
Harald Schneider
Tel: +44 20 79426058
Email: h.schneider@nhm.ac.uk

Received: *6 October 2015*
Accepted: *16 November 2015*

New Phytologist (2016) **210:** 1072–1082
doi: 10.1111/nph.13833

Key words: chromosome number, chromosome structure, DNA C-values, genome size, macroevolution, phylogeny, polyploidy, pteridophytes.

Summary

- The genome evolution of ferns has been considered to be relatively static compared with angiosperms. In this study, we analyse genome size data and chromosome numbers in a phylogenetic framework to explore three hypotheses: the correlation of genome size and chromosome number, the origin of modern ferns from ancestors with high chromosome numbers, and the occurrence of several whole-genome duplications during the evolution of ferns.
- To achieve this, we generated new genome size data, increasing the percentage of fern species with genome sizes estimated to 2.8% of extant diversity, and ensuring a comprehensive phylogenetic coverage including at least three species from each fern order.
- Genome size was correlated with chromosome number across all ferns despite some substantial variation in both traits. We observed a trend towards conservation of the amount of DNA per chromosome, although Osmundaceae and Psilotaceae have substantially larger chromosomes. Reconstruction of the ancestral genome traits suggested that the earliest ferns were already characterized by possessing high chromosome numbers and that the earliest divergences in ferns were correlated with substantial karyological changes.
- Evidence for repeated whole-genome duplications was found across the phylogeny. Fern genomes tend to evolve slowly, albeit genome rearrangements occur in some clades.

Introduction

Ever since the first comprehensive studies on the genetics of ferns, the evolution of fern genomes has been considered paradoxical owing to the conservation of high chromosome numbers in taxa with demonstrated diploid gene expression (Haufler, 1987, 2002, 2014). This paradox led to the hypothesis that ferns underwent multiple cycles of polyploidy (whole-genome duplications (WGDs)) accompanied by subsequent diploidization involving gene silencing, but without apparent chromosome loss, so high chromosome numbers were retained (Haufler, 2002, 2014). Support for this hypothesis has been provided by observations that polyploidy contributes to *c.* 31% of speciation events in ferns compared with *c.* 15% in angiosperms (Wood *et al.*, 2009). Recurrent WGD events without subsequent reduction in chromosome number and genome size may hence explain several characteristics of fern genomes, including the 80-fold variation in chromosome number (ranging from $2n = 18$ in *Salvinia natans* to $2n = 1440$ in *Ophioglossum reticulatum*), 94-fold genome size variation in ferns (ranging from 1C = 0.77 pg in *Azolla microphylla* to 1C = 72.68 pg in *Psilotum nudum*), the highest chromosome number of any plant known to date ($2n = 1440$), and the average chromosome number in homosporous ferns ($n = 57.05$), greatly exceeding the average chromosome number in angiosperms ($n = 15.99$) (Klekowski & Baker, 1966; Leitch & Leitch, 2012, 2013; Barker, 2013; Henry *et al.*, 2015) (the estimated average chromosome number for ferns is updated later). Haufler's hypothesis predicted a correlation between genome size and chromosome number that has so far been found in ferns but not in angiosperms or gymnosperms (Nakazato *et al.*, 2008; Bainard *et al.*, 2011; Barker, 2013; Leitch & Leitch, 2013).

Based on these observations, fern genomes are considered to have shown greater stability in their chromosome structure over the last 400 million yr compared with the sister lineage of ferns – the seed plants (Leitch & Leitch, 2012; Haufler, 2014). This hypothesis appears to be consistent with the recent proposal of

*These authors contributed equally to this work.

'static genomes' in the royal ferns (Bomfleur *et al.*, 2014; Schneider *et al.*, 2015) and arguments suggesting that chromosome size expansion via accumulation of repeats plays only a minor role in the evolution of fern genomes compared with seed plants (Wagner & Wagner, 1980; but see Dyer *et al.*, 2013). The prediction is also consistent with the relatively small number of studies providing evidence for single chromosome gains and losses in ferns (Lovis, 1977), including recent studies that have incorporated phylogenetic evidence in their analyses, for example *Hymenophyllum* (Hennequin *et al.*, 2010), *Lepisorus* (Wang *et al.*, 2010) and the *Loxoscaphe* complex in *Asplenium* (Bellefroid *et al.*, 2010).

The hypothesis of recurrent cycles of hybridization, WGD and conservation of chromosomes has been challenged owing to the lack of strong evidence in expression sequence tag (EST) data in the polypod ferns *Ceratopteris richardii* and *Pteridium aquilinum* (Nakazato *et al.*, 2008; Barker & Wolf, 2010; Barker, 2013). Nevertheless, recent transcriptome data for the horsetail *Equisetum giganteum* is not inconsistent with multiple WGD events contributing to the high chromosome numbers in this species ($2n = 216$; Vanneste *et al.*, 2015).

It is possible to argue that artefacts created by sampling biases, for example, low taxonomic coverage of existing genome size measurements, could mislead some of these interpretations. In particular, the critical observation of the correlation between genome size and chromosome number in ferns may be affected by the current taxonomic sampling comprising < 1% of extant ferns compared with other land plants, especially angiosperms (1.8% coverage according to Leitch & Leitch, 2013), and the absence of genome size data for some phylogenetically important clades of ferns. To address this issue, we designed a study to test several of the major predictions derived from the existing hypotheses, including the repeated cycles of polyploidization, high chromosome numbers in the ancestors of extant ferns, and the conservation of chromosome size (Soltis & Soltis, 1987; Haufler, 2002, 2014) by increasing the number of reliable genome size measurements to cover > 2.5% of the taxonomic diversity of ferns. This included at least two species from each of the 11 orders of ferns currently recognized (Smith *et al.*, 2006; Christenhusz & Chase, 2014), plus a comprehensive sampling of the tree fern lineage (Cyatheales) and basal polypod (Polypodiales) families (e.g. Dennstaedtiaceae, Lindsaeaceae, Lonchitidaceae, Saccolomataceae). To achieve this, we generated new genome size data using best-practice techniques and combined these with existing measurements available in the Pteridophyte DNA C-values database (Bennett & Leitch, 2012). We also assembled a comprehensive database of all chromosome numbers published for ferns. This strategy provided us with the evidence to test the outlined hypotheses using a phylogenetic framework based on our current understanding of the fern phylogeny (Wickett *et al.*, 2014). The analyses confirm the predicted positive correlation between chromosome number and genome size in ferns, demonstrate that the main lineages of extant ferns originated from ancestors having already high chromosome numbers and a chromosome size that is generally conserved through the phylogeny of ferns, and show repeated cycles of WGD and karyotypic changes, which have affected not only the early diverging fern clades but also the more derived lineages of leptosporangiate ferns.

Materials and Methods

Obtaining new genome size data

New genome size measurements were made using freshly collected specimens from the living collections of three institutes: Fairy Lake Botanical Garden at Shenzhen (China), The Royal Botanic Gardens, Kew (UK), and The Royal Botanic Garden Edinburgh (UK) together with additional samples from the private collection of M. Christenhusz, Kingston-upon-Thames (UK). Vouchers were deposited at the corresponding herbaria of these institutes (PI (Institute of Botany of the Chinese Academy of Science, Beijing, China), K (Royal Botanic Gardens, Kew, UK), E (Royal Botanic Gardens Edinburgh, Edinburgh, UK)). To obtain genome size measurements, we selected the most appropriate of four available reference standards for each fern species: *Petroselinum crispum* 'Champion Moss Curled' (2C = 4.50 pg; Obermayer *et al.*, 2002), *Pisum sativum* 'Ctirad' (2C = 9.09 pg; Doležel *et al.*, 1992), *Vicia faba* 'Inovec' (2C = 26.90 pg; Doležel *et al.*, 1992), and *Allium cepa* 'Ailsa Craig'. The genome size for *Allium cepa* was recalibrated using both *Vicia faba* and *Pisum sativum*, and a value of 2C = 34.89 pg was determined, which is larger than the 2C-value of 33.55 pg (Bennett *et al.*, 2000) but identical to that of Doležel *et al.* (1998). Based on the peak quality in the flow histograms, one out of four nuclei isolation buffers (Doležel *et al.*, 1989; Ebihara *et al.*, 2005; Loureiro *et al.*, 2007; Supporting Information Table S1) was selected. The nuclear DNA content was measured using the one-step procedure (Doležel *et al.*, 2007) involving the co-chopping of a leaf fragment of the fern with leaf material of the selected reference standard directly into 2 ml of the chosen buffer in a Petri dish over ice, filtering and staining with propidium iodide. For each sample, *c.* 5000 particles were measured using a PAII or CyFlowSL Partec flow cytometer (Partec GmbH, Goettingen, Germany) fitted with a 100 W high-pressure mercury lamp or a 100 mW green (532 nm solid-state Cobalt Samba laser; Cobolt AB, Solna, Sweden), respectively. The resulting flow histograms were analysed using the Partec software for flow cytometry FloMax 2.7. Where possible, nine measurements were made per sample. Mean 2C-values and standard deviations were calculated for each taxa based on the ratio of fluorescence between the fern and calibration standard (Pellicer & Leitch, 2014).

As extant ferns, especially homosporous species, often have multiple cytotypes (e.g. *Ophioglossum reticulatum* has individuals with $2n = 240, 480, 720, 960, 1440$; Khandelwal, 1990) which differ in holoploid 2C-values because of differences in the number of monoploid chromosome sets, insights into the genome size stability of the monoploid chromosome set can be obtained by calculating the monoploid genome size (= 1C*x*-value; see Greilhuber *et al.*, 2005). To do this requires identification of the base chromosome number *x*, which has been recognized as a challenge

in ferns (Duncan & Smith, 1978). Here, we accepted the lowest gametophytic chromosome number (n) reported for a genus (including any counts that could be explained by single chromosome gains and losses) to be the most likely x for the investigated genus and, where possible, family (Manton & Vida, 1968). In many cases, chromosome counts were not available for the actual individual used for genome size estimation, so there is the potential that the 1Cx-value is over- or underestimated if an incorrect n or x value is assumed. To take this into account, several approaches were taken: wherever possible, chromosome counts were made for the plants studied; chromosome counts/ploidy levels were assumed based on previous reports in the literature; genome sizes were estimated for the same individual which had previously been reported by other authors; and the modelling of genome size and chromosome evolution (see later) were conducted multiple times using different x values to investigate the effect of this uncertainly on the reconstruction of ancestral states. In addition, to avoid the ambiguity introduced by the selection of x, we also calculated a further value as a proxy for the average chromosome size of a species. This was determined by dividing the 2C-value by the somatic chromosome number $2n$. This provides an insight into the extent of variation in the average chromosome size across ferns. Nevertheless, it is also recognized that this value may be incorrectly calculated for species where an actual chromosome count was not made for the individual used for genome size estimation. The same approaches were used as earlier to take into account these uncertainties and hence increase the robustness of the analyses.

Integration of new and existing data

We assembled the following data for each corresponding taxon: 1C-values, chromosome number ($2n$) and plastid DNA sequences (see Table S1). We checked and assembled all published 1C-values of ferns using the Pteridophyte DNA C-values database (Bennett & Leitch, 2012) or recent publications (Ekrt *et al.*, 2009; Bainard *et al.*, 2011; Nitta *et al.*, 2011; Williams & Waller, 2012; Bou Dagher-Kharrat *et al.*, 2013; Chang *et al.*, 2013; Dyer *et al.*, 2013; Pustahija *et al.*, 2013). The chromosome numbers of ferns were obtained from online databases (http://ccdb.tau.ac.il and http://chromosomes.binoz.uj.edu.pl/chromosomes in August 2015, plus the unpublished database comprising all fern chromosome counts assembled by H.S.). To estimate taxonomic coverage, we accepted the estimates of species diversity published in Smith *et al.* (2006).

Phylogenetic framework

The phylogenetic framework was generated by assembling a data matrix of at least one of three plastid genome regions (*atpA*, *atpB*, *rbcL*) accessible in GenBank (Table S1) for taxa with genome size data available. In some cases, DNA sequences of closely related species were used, as DNA sequences were not available for the sampled species.

The molecular dataset was assembled and aligned using Mesquite v.2.75 (Maddison & Maddison, 2014) and the best-fitting model of molecular evolution was determined using jModeltest v.2.1.4 (Darriba *et al.*, 2012) to generate the selection statistics and the Bayesian information criterion. The phylogenetic framework was then estimated using MrBayes v.3.2 (Ronquist *et al.*, 2012) with two independent runs for three million generations with the convergence verified using Tracer v.1.6 (Rambaut *et al.*, 2014). To test the impact of the uncertain phylogenetic placement of horsetails, we also used alternative phylogenetic hypotheses including Equisetales sister to all other ferns, Equisetales sister to the clade comprising Ophioglossales and Psilotales, Equisetales sister to Marattiales, and Equisetales sister to Polypodiales (Pryer *et al.*, 2001; Kim *et al.*, 2014; Wickett *et al.*, 2014). When required, polytomies were artificially resolved according to current consensus on fern phylogeny (Lehtonen, 2011). The phylogenetic signal of chromosome numbers ($2n$), base number (x), holoploid genome size (1C), monoploid genome size (1Cx), and the average chromosome size (2C/$2n$) was analysed with the traits module in Phylocom v.4.2 (Webb *et al.*, 2008) with significance tested via randomization of trait values across the Bayesian trees. Furthermore, the phylogenetic signal of each trait was explored by applying phylogenetic dependent tests alongside phylogenetic independent tests, such as Spearman's rank correlation as implemented in the Hmisc package in R 3.14 (Harrell, 2014). Phylogenetic dependent tests comprised the calculation of phylogenetic independent contrasts (PICs; Felsenstein, 1985) and Felsenstein's contrast correlation between positivized contrasts (Garland *et al.*, 1993) using the Pdap:pdtree package 1.16 (Midford *et al.*, 2011) in Mesquite 2.75 with the trees obtained from the Bayesian analyses. Ancestral character states of holoploid genome size (a1C), monoploid genome size (a1Cx), and average chromosome size (a2C/$2n$) were estimated using BayesTraits v.2.0 (Pagel *et al.*, 2004). The latter analyses were performed with 100 trees randomly drawn from the posterior distribution of the phylogenetic reconstruction and values were estimated from a Bayesian Markov chain Monte Carlo analysis of two million generations sampling every 1000th generation and discarding the first 10 000 generations. Reconstructed trait values were averaged over the posterior distribution with a 95% confidence interval calculated. The base chromosome number was inferred using ancestral state reconstruction in ChromEvol v.2.0 (Mayrose *et al.*, 2010). The ancestral diploid chromosome number (a2n) was estimated using a guide tree derived from the Bayesian consensus tree with taxa with uncertain chromosome number set to an equal likelihood. Character states were optimized using the model assuming constant rate of chromosome gain, loss and duplication along with an estimated rate of demiduplication because this model was selected based on the output of the initial analyses with 10 models of chromosome evolution with the Akaike information criterion (Table S2).

Results

The taxonomic coverage of ferns with genome size data was increased to 2.8% by the addition of genome size measurements for 110 newly studied species, corresponding to first

Table 1 Summary of the mean, minimum (Min), maximum (Max) and range (Max/Min) of genome size (1C) and chromosome number (2*n*) data available for all ferns and for each of the 11 recognized orders of ferns

	SN*	SN-1C	TC-1C (%)	Mean 1C (pg)	Min 1C (pg)	Max 1C (pg)	*x*-fold-1C	SN-2*n*	TC-2*n* (%)	Mean 2*n*	Min 2*n*	Max 2*n*	*x*-fold-2*n*
All ferns	9118	208	2.3	14.29	0.77	72.03	93.54	2639	28.9	121.0	18	1440	80.0
Equisetales	14	15	93.3	21.32	12.78	30.35	2.37	15	100	216	216	216	0.0
Ophioglossales	80	12	15.0	28.35	10.22	65.55	6.41	65	81.2	312.1	88	1440	16.4
Psilotales	12	2	16.7	72.35	72.03	72.68	1.18	12	100	269.7	104	416	4.0
Marattiales	150	8	5.3	10.66	6.9	13.99	2.03	20	1.3	101.6	78	160	2.0
Osmundales	20	9	45.0	15.68	13.46	21.01	1.56	14	70.0	44	44	44	0.0
Hymenophyllales	600	6	1.0	17.61	10.73	21.31	1.99	162	27.0	79.5	22	256	11.6
Gleicheniales	140	3	2.1	2.96	2.43	3.26	1.34	24	17.1	84.9	40	232	5.8
Schizaeales	155	7	4.5	13.74	6.16	22.6	3.67	50	32.2	190.8	56	1080	19.3
Salviniales	91	3	3.3	2.38	0.77	4.08	5.30	25	27.5	58.4	18	120	6.7
Cyatheales	663	12	1.8	7.91	2.52	12.57	4.99	85	12.8	151.6	92	276	3.0
Polypodiales	7192	126	1.8	12.19	3.39	60.50	17.84	2166	34.3	114.1	22	576	26.2

*Total species numbers based on Smith *et al.* (2006).
SN, species number; TC, taxon coverage.

records for 47 genera, and one order (Gleicheniales) (Tables 1, S1). Each order of ferns now includes genome data for at least two distinct species, with taxon coverage ranging from 1.0% for Hymenophyllales to 93.3% for Equisetales. The data also include the first measurements for crucial genera such as *Tmesipteris*, the sister genus of *Psilotum* (Table S1), and three genera belonging to the early diverging lineage of the species-rich order Polypodiales, that is, *Lindsaea*, *Lonchitis* and *Saccoloma*. The new measurement of *Tmesipteris* (2C = 144.1 pg) is similar to the extremely large genome previously recorded in the whisk fern *Psilotum nudum* (2C = 145.2 pg), which has the largest genome size of any fern so far studied (Tables 1, S1). Nevertheless, our new data show that large genomes are not restricted to early-diverging ferns, as shown by the discovery of considerably larger genomes in some Polypodiales than previously reported. These were found in species nested in a clade within eupolypods I, that is, *Dracoglossum plantagineum* (2C = 85.5 pg) and *Mickelia nicotianifolia* (2C = 121.0 pg), but also within the early diverging lineages of Polypodiales, that is, *Saccoloma domingense* (2C = 77.5 pg). Relatively large genomes were also recorded in *Bolbitis* (2C = 50.7 and 55.3 pg) and *Elaphoglossum* (2C up to 67.0 pg), which form a clade together with *Mickelia*. Thus large genomes may be restricted to a few clades in the derived ferns.

The smallest mean genome sizes were found in the leptosporangiate orders Gleicheniales and Salviniales, the latter including the smallest genome size of any fern so far reported (i.e. Azolla microphylla; 1C = 0.77 pg, Obermayer *et al.*, 2002). However, whether these small mean values reflect the limited sample sizes (with only three estimates for each order) remains to be determined. Certainly the more extensive range of chromosome numbers encountered in each group (Salviniales $2n = 18–120$; Gleicheniales $2n = 40–232$) suggests that larger genome sizes may be encountered in these orders as data increase.

Based on the collation of chromosome data from different sources, chromosome counts are now available for 2639 species corresponding to 28.9% of all ferns, with numbers in the range $2n = 18–1440$, and an estimated mean gametic chromosome number of $n = 60.5$ for all ferns included (Table 1; according to the database incorporating all accessible chromosome counts: $n = 63.5$ for all ferns, $n = 63.5$ for homosporous ferns and $n = 58.4$ for heterosporous ferns). The 11 orders of ferns show distinct differences in the extent of variation in genome size and chromosome number (Table 1; Fig. 1).

Overall there was a significant positive relationship between homoploid genome size (1C) and chromosome number (Table 2; Fig. 2). Both phylogenetic and nonphylogenetic tests found significant support for a correlation between monoploid genome size (1C*x*) and estimates of base chromosome number ($P < 0.001$; Table S1). However, it is notable that, based on available data, there was no clear relationship between the range in chromosome sizes and genome size across the phylogeny of ferns ($P < 0.01$ for range-1C/range-2*n*; Table 1). For example, while the constant chromosome number of $2n = 44$ coincides with a low variation in genome size (1.56-fold; 1C = 13.46–21.01 pg) in Osmundales, genome sizes in Equisetales range over 2.37-fold (1C = 12.78–30.35 pg) and yet their chromosome number is also highly conserved, with all species analysed to date having $2n = 216$ (Fig. 1; Table 1).

Discussion

Prediction 1: There is a positive correlation between chromosome number and genome size

The considerably increased sampling of genome size and chromosome data from 231 taxa across the phylogenetic tree of ferns, analysed using either the phylogeny-independent Spearman's correlation coefficient (SCC) or the phylogeny-dependent Felsenstein contrast correlation (FCC) (Table 2) clearly supports the hypothesis that genome size and chromosome number are correlated in ferns. The correlation was also supported for most of the major subgroups when analysed separately. The only exceptions were found within the

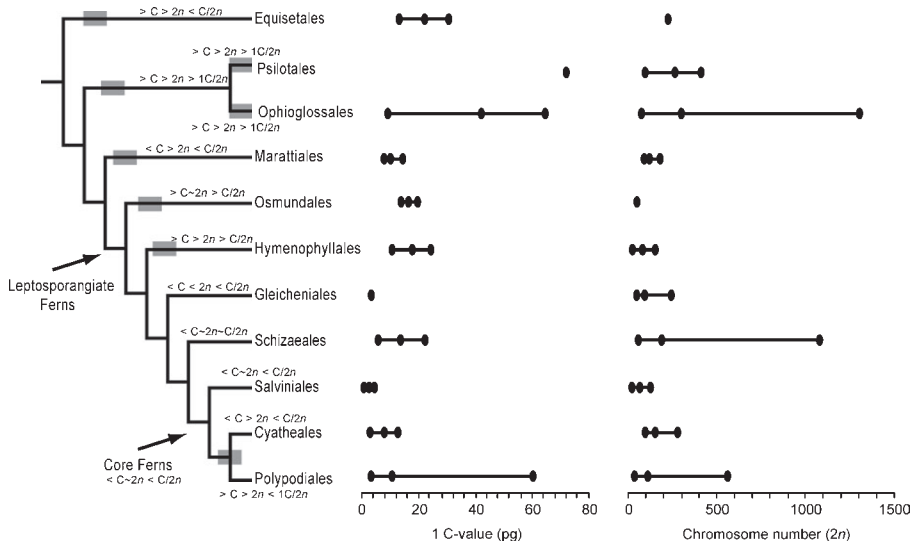


Fig. 1 Summary of the observed distribution of genome size and chromosome number variation among the 11 orders of ferns. Bars indicate the range of variation with the maximum, mean and minimum values indicated via dots. The phylogeny summarizes the currently accepted hypothesis with the horsetails considered sister to the remaining fern clade (for alternative topologies, see Supporting Information Fig. S2). Terminal taxa correspond to orders according to Smith *et al.* (2006) and Christenhusz *et al.* (2011), whereas leptosporangiate ferns and core ferns are marked by arrows. Based on the data given in Tables 1 and 3, some clades show notable trends such as increase (>)/decrease (<) of genome size (1C), increase of chromosome numbers (2*n*), or changes in chromosome size (2C/2*n*). Grey squares indicate clades that show some evidence for whole-genome duplications.

Table 2 Summary of the test results for the prediction of a positive correlation between genome size (1C, pg) and chromosome number (2*n*) across major clades recognized in the current fern phylogeny (Pryer *et al.*, 2004; Smith *et al.*, 2006; Lehtonen, 2011) in both a phylogeny-dependent (Spearman's Correlation Coefficient) and phylogeny-independent (Felsenstein Contrast Correlation) context

		Nonphylogenetic test Spearman's correlation coefficient		Phylogenetic test Felsenstein contrast correlation		
	No. taxa	ρ	*P*-value	No. contrasts	*r*	*P*-value
All ferns	185	0.44	< 0.001	166	0.50	< 0.001
Nonleptosporangiate fern grade	38	0.61	< 0.001	27	0.27	0.16
Leptosporangiate ferns	147	0.22	0.006	138	0.61	< 0.001
Leptosporangiate ferns without Osmundales	136	0.42	< 0.001	129	0.61	< 0.001
Core leptosporangiate ferns	125	0.44	< 0.001	119	0.60	< 0.001
Polypodiales	111	0.46	< 0.001	104	0.60	< 0.001
Eupolypod ferns	90	0.44	< 0.001	86	0.61	< 0.001
Eupolypods I	38	0.39	0.02	34	0.54	< 0.001
Eupolypods II	52	0.40	0.003	51	0.69	< 0.001

nonleptosporangiate fern grade and the eupolypods I (Polypodiaceae s.l.) clade where FCC failed to support a positive correlation, even though SCC did (Table 2). Thus, while the increase of chromosome numbers, through either polyploidy or other chromosomal processes, does not seem to be accompanied by extensive loss of DNA in the majority of fern lineages (a situation that contrasts strongly with observations in angiosperms; Leitch & Bennett, 2004), this configuration is not universal amongst ferns. However, the recovery of the same correlation between the monoploid genome size 1C*x* and the base chromosome number *x* ($P < 0.001$) rejects the argument that this pattern is caused by neopolyploidy alone. Future work is certainly needed to focus within the nonleptosporangiate fern grade and eupolypods I to identify whether the lack of support for the correlation using FCC is a sampling artefact, or the result of genuine distinctive genomic processes operating within these lineages as suspected (see Prediction 3 later).

Prediction 2: The main lineages of extant ferns originated from ancestors with high chromosome numbers and chromosome size is broadly conserved throughout the phylogeny of ferns

The results presented here reconstruct high ancestral chromosome numbers (a2*n*) for each of the main extant lineages of ferns, with estimated a*2n* ranging from 40 in Schizaeales to 216 in Equisetales and an estimate of a2*n* = 44 for the ancestors of all

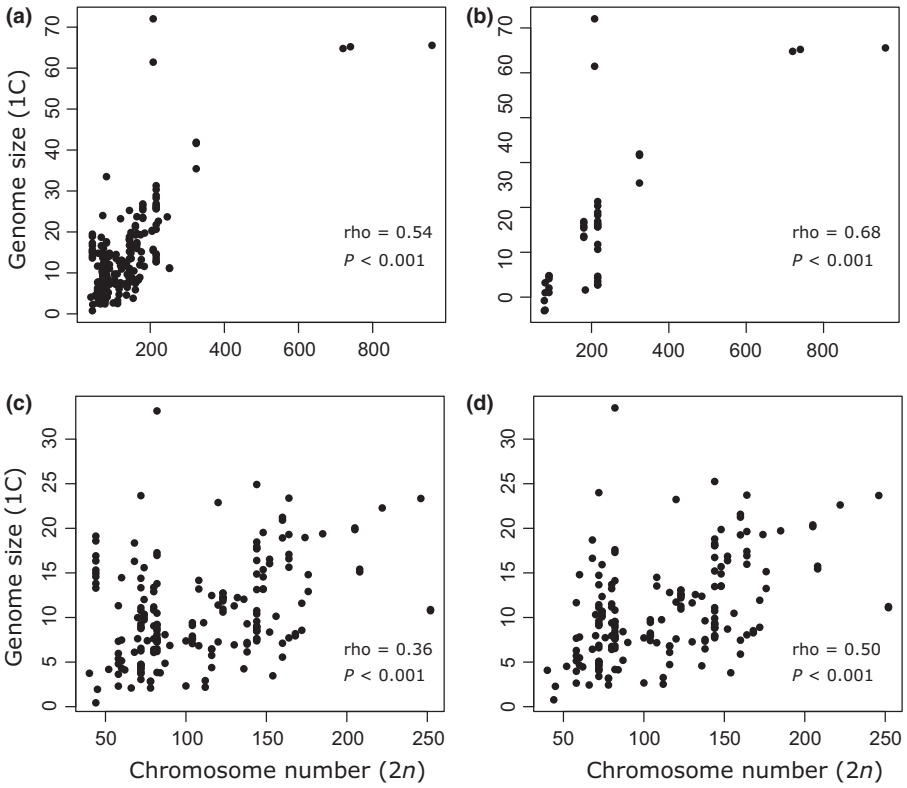


Fig. 2 Correlation plots of chromosome number (2*n*; *x*-axis) vs genome size (2C; *y*-axis). Each dot corresponds to one accession. Linear regression statistics are given by rho and *P*-values. (a) Plot including all ferns; (b) plot including all nonleptosporangiate fern orders; (c) plot including all leptosporangiate fern orders; (d) plot including leptosporangiate fern orders without Osmundales.

ferns (Table 3). Given the high a2*n* of nonleptosporangiate ferns (a2*n* = 44–216), together with the estimated a2*n* = 48 for leptosporangiate ferns (both with and without Osmundales; Table 3), the data suggest that the main lineages of extant ferns originated from ancestors having already high chromosome numbers (Table 3). The presence of even higher a2*n* in some of the derived lineages of leptosporangiate ferns such as Hymenophyllales (a2*n* = 72), Cyatheales (a2*n* = 68) and Polypodiales (a2*n* = 64), especially all eupolypods I (a2*n* = 82), suggest that further increases have taken place along their evolutionary history leading to the establishment of even higher chromosome numbers at the base of these lineages (see Prediction 3 later).

With the exception of Marattiales, all nonleptosporangiate fern lineages are characterized by possessing large genome sizes with a mean 1C-value > 20.0 pg (Table 1) and an ancestral monoploid genome size (a1C*x*) > 11.0 pg (Table 3).

Extremely large genomes with 1C > 35.0 pg (Table 1) were concentrated in the sister orders Ophioglossales and Psilotales, but were also found for the first time in genera of derived leptosporangiate ferns within Polypodiales (i.e. *Dracoglossum*, *Mickelia* and *Saccoloma*; see Table S1). Such data, together with the high a2*n* estimates, support the predicted repeated establishment of ferns with large genomes through polyploidization without subsequent genome downsizing in these lineages. However, denser taxonomic sampling may show that this general trend is not strictly conserved. In particular, three patterns may require special attention: lineages showing evidence for reduction of chromosome number (e.g. Hymenophyllales); increased chromosome number without substantial increase of genome size (e.g. Cyatheales); and frequent variation of the chromosome number unlinked with ploidy (e.g. clade comprising relatives of *Blechnum, Onoclea* and *Woodwardia*).

It has often been stated that fern chromosomes are generally characterized by being small and uniform in size, in contrast to the diversity encountered in seed plants (e.g. Wagner & Wagner, 1980; Nakazato *et al.*, 2008). Indeed, based on available data, an analysis of the range of chromosome sizes (2*n*/2C) shows that while angiosperm chromosome sizes range over 3100-fold (2*n*/2C, 0.003–9.300 pg per chromosome), in ferns they range only 31-fold, from 0.035 to 0.955 pg per chromosome. The results presented here, in which an estimate of ancestral chromosome sizes have been reconstructed using a2C/2*n*, broadly support this predicted overall conservation of small chromosomes throughout fern evolution, with an inferred a2C/2*n* value for eight out of 11 fern orders falling within the standard deviation of the mean value (Table 3). Nevertheless, there are exceptions, with larger chromosomes in the two previously mentioned sister orders, Ophioglossales (a2C/2*n* = 0.270 pg) and Psilotales (a2C/2*n* = 0.621 pg); the earliest diverging order of leptosporangiate ferns, the royal ferns (Osmundales, a2C/2*n* = 0.690 pg); the clade comprising *Blechnum*, *Onoclea*, *Woodwardia* and relatives; and the clade comprising relatives of *Elaphoglossum* (Polypodiales subfam. Dryopteridaceae) (Tables 3, S1).

Within the leptosporangiate ferns, the uniqueness of the chromosomes of Osmundales has previously been recognized (Manton, 1950), and while the remaining lineages show evidence for rather limited chromosome size variation, as noted earlier, some clades do show distinct trends that deviate from this. For example, a trend towards smaller genomes and chromosomes in the heterosporous Salviniales (a1C*x* = 4.81 pg and

New Phytologist

Table 3 Summary of the reconstructed ancestral character states inferred from the obtained genome size measurements and chromosome numbers. The ancestral values were reconstructed with either BayesTraits 2.0, that is, a1C, a1C*x*, and a2C/2*n*, over 100 trees drawn from the posterior distribution of the Bayesian phylogenetic reconstruction shown in Supporting Information Fig. S2, or ChromEvol, that is, a2*n*, with implementation of the estimated parameter values of the model of chromosome evolution for all ferns and basal lineages only: gain = 1.37/2.14; loss = 2.97/2.02e–10; duplication = 1.08/1.25; demiduplication = 0.95/0.24 with the model likelihood of −438.5/−62.7, respectively

	a1C (pg)	a1C*x* (pg)	a2*n*	a2C/2*n*
All ferns	28.67 ± 0.32	14.60 ± 0.08	44	0.449 ± 0.005
Psilotales + Ophioglossales	37.40 ± 0.34	18.84 ± 0.19	44	0.523 ± 0.005
Psilotales	67.88 ± 0.25	34.27 ± 0.14	104	0.621 ± 0.003
Ophioglossales	32.19 ± 0.34	11.59 ± 0.20	90	0.270 ± 0.004
Equisetales	19.63 ± 0.20	19.80 ± 0.11	216	0.187 ± 0.003
Marattiales	10.61 ± 0.19	9.62 ± 0.12	78	0.233 ± 0.003
Leptosporangiate ferns	15.80 ± 0.33	12.78 ± 0.21	48	0.424 ± 0.004
Osmundales	15.91 ± 0.18	15.44 ± 0.10	44	0.690 ± 0.002
Leptosporangiate ferns (minus Osmundales)	11.84 ± 0.31	10.04 ± 0.18	48	0.266 ± 0.005
Hymenophyllales	14.97 ± 0.29	14.04 ± 0.17	72	0.327 ± 0.005
Gleicheniales	10.24 ± 0.30	9.58 ± 0.17	48	0.250 ± 0.005
Schizaeales	11.60 ± 0.44	9.16 ± 0.25	40	0.251 ± 0.006
Core leptosporangiate ferns	9.67 ± 0.26	8.73 ± 0.16	46	0.142 ± 0.004
Salviniales	5.68 ± 0.35	4.81 ± 0.21	46	0.099 ± 0.005
Cyatheales	9.45 ± 0.16	8.91 ± 0.10	68	0.134 ± 0.002
Polypodiales	16.07 ± 0.30	7.34 ± 0.18	64	0.156 ± 0.004
Dennstaedtiaceae	10.78 ± 0.24	5.70 ± 0.14	58	0.159 ± 0.004
Pteridaceae	9.13 ± 0.25	5.13 ± 0.15	60	0.177 ± 0.003
Eupolypods	10.95 ± 0.19	7.44 ± 0.11	54	0.190 ± 0.002
Eupolypods I	13.24 ± 0.18	9.11 ± 0.10	82	0.228 ± 0.003
Dryopteridaceae	13.25 ± 0.18	9.11 ± 0.10	82	0.229 ± 0.002
Nephrolepidaceae	8.69 ± 0.12	8.90 ± 0.05	82	0.216 ± 0.002
Tectariaceae	12.02 ± 0.16	11.56 ± 0.09	82	0.281 ± 0.002
Davalliaceae	10.60 ± 0.15	10.24 ± 0.09	80	0.257 ± 0.002
Polypodiaceae	12.02 ± 0.14	10.86 ± 0.07	74	0.297 ± 0.002
Eupolypods II	9.90 ± 0.15	7.05 ± 0.08	54	0.179 ± 0.002
Cystopteridaceae	8.00 ± 0.17	4.98 ± 0.09	82	0.132 ± 0.003
Thelypteridaceae	10.09 ± 0.15	7.12 ± 0.09	50	0.195 ± 0.003
Onocleaceae	13.29 ± 0.16	12.78 ± 0.10	78	0.329 ± 0.003
Blechnaceae	12.06 ± 0.15	10.93 ± 0.09	70	0.306 ± 0.003
Athyriaceae	12.02 ± 0.13	8.61 ± 0.08	80	0.223 ± 0.002
Woodsiaceae	7.40 ± 0.17	5.33 ± 0.10	78	0.142 ± 0.003
Aspleniaceae	8.86 ± 0.16	5.97 ± 0.09	72	0.173 ± 0.002

a2C/2*n* = 0.099 pg; Table 3) is detected, while a tendency towards larger genomes coinciding with larger chromosomes in Hymenophyllales (a1C*x* = 14.04 pg and a2C/2*n* = 0.327 pg) and several clades of eupolypods I and II is also evident (Table S1).

In summary, the recovered results are consistent with the prediction that ancestors of the extant lineages of ferns shared high chromosome numbers, whereas the previously predicted stability in chromosome size is found in some, but not all, of the main fern lineages.

Prediction 3: Repeated cycles of WGD and chromosome composition changes in the ancestors of ferns contribute to the diversity and high chromosome numbers encountered in extant ferns

Recurrent WGDs may explain the observed accumulation of large genomes coinciding with high chromosome numbers found in early diverging lineages such as Equisetales (a2*n* = 216), Marattiales (a2*n* = 78), Ophioglossales (a2*n* = 90) and Psilotales (a2*n* = 104), based on an estimated a2*n* of 44 for all ferns and a2*n* of 44 for the common ancestor of the Psilotales + Ophioglossales clade. These results are consistent with the evidence for a WGD in the common ancestor of the extant horsetails detected in transcriptome data (Vanneste *et al.*, 2015).

Based on an estimated a2*n* = 48 reconstructed at the base of all leptosporangiate ferns, there is also evidence for at least three independent putative WGDs within this species-rich fern clade; namely the ancestors of Hymenophyllales (filmy ferns; a2*n* = 72), the base of the sister lineages Cyatheales (tree ferns; a2*n* = 68) and Polypodiales (a2*n* = 64), and the ancestors of eupolypods I (a2*n* = 82) (see Fig. 1).

In Hymenophyllales, the extant crown group is predicted to have originated from ancestors with an inferred a2*n* = 72, and yet some extant species belonging to the derived *Hymenophyllum* clade are reported to have chromosome numbers as low as 2*n* = 22. It is suggested that these are likely to be the result of chromosome deletions following WGD (see Hennequin *et al.*, 2010). Support for a WGD at the base of Hymenophyllales is also suggested from the ancestral monoploid genome size (a1C*x*)

of 14.04 pg, which is larger than the ancestral a1C*x* of the other two precore leptosporangiate orders, Gleicheniales (a1C*x* = 9.58 pg) and Schizaeales (a1C*x* = 9.16 pg), and leptosporangiate ferns minus Osmundales clade (a1C*x* = 10.04 pg). Nevertheless, the observed 1C-values estimated for the three Gleicheniales species studied (i.e. 1C = 4.9, 6.4, 6.5 pg, Table S1) suggest that an even lower a1C-value, closer to half of the current estimate of a1C = 9.58 pg may be more realistic for this order.

While the higher a2*n* reconstructed for Cyatheales (a2*n* = 68) and Polypodiales (a2*n* = 64) suggest a WGD in their most recent common ancestor, the increase in the a1C to 16.07 pg at the base of Polypodiales compared with a1C = 9.67 pg of core leptosporangiates and a1C = 9.45 pg of Cyatheales (Table 3) suggests that if a WGD did give rise to the higher a2*n* observed in Cyatheales, this may have been accompanied by some genome downsizing.

In this context, it is important also to note that core leptosporangiate ferns and their three orders (Salviniales, Cyatheales and Polypodiales) tend to show lower inferred a2C/2*n* values than other fern orders (Table 3), perhaps suggesting some reduction in genome size at the base of this lineage. However, such a pattern needs to be considered in the context of the observed variation of these values among Polypodiales, for example, increased 2C/2*n* values in some eupolypods. For example, the clade of eupolypods II comprising the relatives of *Blechnum*, *Onoclea* and *Woodwardia* shows evidence of karyological changes, as indicated by the a2C/2*n* values of 0.329 and 0.306 pg, respectively, compared with values found in the closely related clade, which all have 2C/2*n* < 0.230 pg (Table 3). Nevertheless, this clade does not show a correlation between chromosome number and genome size ($P = 0.118$). Clearly further sampling within the core leptosporangiates is required to even out the effects of limited sample size that may have resulted in a too low estimate of the a2*n* in this major clade of derived ferns a2*n* = 54 (Table 3).

In the eupolypods I, the clade comprising the two mainly epiphytic lineages, Davalliaceae/Davallioideae and Polypodiaceae s.s. Polypodioideae, plus its relatives the climbing Oleandraceae/Oleandroideae and mainly terrestrial Tectariaceae/Tectarioideae, show increased a2C/2*n* values compared with the mainly terrestrial Dryopteridaceae/Dryopteridoideae and related clades. These coincide with a reduction in the ancestral chromosome number from a2*n* = 82 (Dryopteridaceae/Dryopteridoideae) via a2*n* = 80 (in Davalliaceae/Davallioideae) to a2*n* = 74 (in Polypodiaceae s.s/Polypodioideae), suggesting chromosome rearrangements following WGD but not accompanied by extensive changes in genome size. By contrast, within Dryopteridaceae/Dryopteridoideae there is evidence of genome size increases in the clade comprising the genera *Bolbitis-Elaphoglossum-Mickelia* (including the largest genome size so far reported for any derived leptosporangiate fern, i.e. 2C = 121.0 pg in *Mickelia nicotianifolia*), whereas other clades show evidence for conservation of genome size. In this context it is worth noting that the estimated ancestral chromosome number is more variable among the clades of eupolypods II (a2*n* = 50–82) than among those of eupolypods I (a2*n* = 74–82). Such results indicate that substantial changes in genome size, chromosome number and chromosome size have taken place during the recent diversification of these derived clades of ferns and point to these genomes being more dynamic than perhaps hitherto recognized.

Overall the results support the prediction that the high chromosome numbers in ferns have resulted, in part, from repeated cycles of WGD and chromosome composition changes that have affected not only the early diverging fern clades but also the more derived lineages of leptosporangiate ferns.

Phylogenetic uncertainty and the robustness of the inferred results

To infer the robustness of the reported results, we explored the impact of alternative phylogenetic hypotheses regarding the relative positions of the five basal orders of ferns (Equisetales, Marattiales, Ophioglossales, Osmundales, Psilotales; Fig. S1). The reported results indicate little variation in the reconstructed genome size estimates caused by changes in the tree topology. For example the reconstructed genome size for the ancestor shared by the sister orders Ophioglossales-Psilotales ranges from a1C = 37.24–38.43 pg across the four topologies explored (Fig. S2). However, we anticipate that the variation for several orders will increase as the amount of genome size data improves, especially in Gleicheniales. Certainly, increased sampling is required to improve our understanding of genome size evolution in Hymenophyllales and the eupolypods. The still relatively low taxon coverage of 2.8% for all ferns, and especially of some species-rich clades such as Hymenophyllales (1.0%) and Polypodiales (2.4%), is a concern, because there is some tantalising evidence that genome reorganization, including genome downsizing and chromosome rearrangements, may have occurred in these lineages, but the data are currently too limited to draw firm conclusions. If such processes have occurred, they will only become recognizable when a denser sampling can provide sufficient sensitivity to detect these patterns.

Other potential limitations of the current data to the accuracy of the analyses include errors arising from genome size estimates, determination of chromosome counts, and violations of the assumption that chromosome number is generally stable in the majority of fern species. Nevertheless, while only a small percentage of specimens used for our genome size measurements had their chromosome numbers determined (< 10%), there was no evidence to support these concerns, because measurements of specifically collected specimens with recently published chromosome counts fitted well into the pattern we observed. It is noted that the effect of sampling density may be especially high in the context of the estimated ancestral diploid chromosome numbers of Psilotales and Polypodiaceae/eupolypods I.

Concluding remarks

In summary, the increased sampling of genome size measurements presented here has provided evidence to support earlier predictions based on hypotheses of repeated WGD events in the ancestors of extant ferns. This includes evidence for repeated WGDs during the diversification of leptosporangiate ferns, for

example Hymenophyllales, Polypodiales and the eupolypods I clade of the derived ferns (Fig. 1). This hypothesis is not inconsistent with the results of investigations based on comparative analyses of the transcriptomes of the polypod ferns *Pteridium* (Dennstaedtiaceae) and *Ceratopteris* (Pteridaceae), which suggested a WGD linked with the divergence of Polypodiales, although the data were difficult to interpret (Barker & Wolf, 2010). Our data also suggest some genomic rearrangements in the common ancestor of Cyatheales and Polypodiales as indicated by the rise of the ancestral chromosome number from around a2n = 46 in core leptosporangiate ferns to a2n = 68 in Cyatheales and a2n = 64 in Polypodiales, and a rise in the ancestral genome size from a1C = 9.67 pg in core leptosporangiates to a1C = 16.07 pg in Polypodiales but not in Cyatheales with an a1C = 9.45 pg. Future studies could concentrate on the recovered evidence for changes in the ancestral chromosome number in Cyatheales and Polypodiales, as this may provide further insights into the nature of the evolutionary processes that have enabled these lineages to diversify in the shadow of angiosperms (Schneider *et al.*, 2004). The limited data for Hymenophyllales suggest that the ancestors for the crown group underwent a WGD, while the low chromosome numbers in some derived *Hymenophyllum* species are a result of substantial chromosome restructuring, probably involving chromosome fusion following WGD. Together with the findings of increased genome sizes in some of the more derived fern clades that include epiphytes, the results may suggest a link between genome size evolution and the colonization of and adaptation to epiphytic habitats.

The new findings are broadly consistent with the generally accepted view that fern genomes are less dynamic than those of angiosperms (Leitch & Leitch, 2013). For example, the 2C/2n – DNA content per chromosome – appears to be broadly conserved throughout the evolution of ferns, with the notable exception of Osmundales and the Psilotales-Ophioglossales clade. Furthermore, our results are consistent with the observations on the fern genome spaces based on low-coverage whole-genome shotgun sequencing of six fern species, especially the absence of a correlation between genome size and repeats in ferns (Wolf *et al.*, 2015). However, this does not mean that all fern genomes are 'frozen' in time. WGD events have clearly played an important role in the evolutionary history of ferns and this has been accompanied by changes in genome size, and in chromosome size and chromosome reorganization in some lineages. Certainly the data hint at substantial divergences in the eupolypod I clade comprising *Bolbitis*, *Dracoglossum*, *Elaphoglossum* and *Mickelia*, highlighting the need for denser sampling within eupolypods as noted earlier. Indeed, it seems possible that as sampling improves, the general trend of a correlation between chromosome number and genome size may not be recovered in some clades of ferns that have undergone recent changes in genome organization.

While an analysis based on chromosome numbers has noted that polyploidy may contribute more frequently to speciation events in ferns compared with angiosperms (see Wood *et al.*, 2009), whether there are fundamental differences in the distribution, frequency and genomic consequences of WGD between ferns and angiosperms will have to wait until there is a sufficient amount of genomic data for ferns to enable meaningful comparisons to be made. Comparing our results with studies on angiosperms (Soltis *et al.*, 2009; Jiao *et al.*, 2011) reveals some striking similarities but also differences such as a lower frequency of WPGs in ferns than in angiosperms. This observation is in conflict with the higher rate of neopolyploidy in ferns compared with angiosperms (Wood *et al.*, 2009). It is certainly felt that at present, the meagre amount of genomic and transcriptome data for ferns, together with the striking differences in species diversity and evolutionary ages between ferns and angiosperms, means that any apparent similarities or differences may be misleading, and at this stage it would be highly speculative to try to establish evolutionary trends solely based on numeric comparisons between these groups. Such discoveries and insights will have to wait until the advances being made in genomic sequencing technologies impact more extensively on the ferns.

Acknowledgements

The authors are grateful to the support by the gardens of RBG Edinburgh, RBG Kew, The Australian National Botanical Garden Canberra, the Botanische Staatssammlung Munich, and colleagues sharing material (S. Siljak-Yakovlev, S. Zona). The project was part of the MSc thesis by J.C. generated in the course 'Taxonomy and Biodiversity' run jointly by the Natural History Museum and Imperial College London. The research was supported by the Chinese Academy of Science senior visiting professorship to H.S., scholarship of the Chinese Scholarship Council to H.L., and BBSRC PhD project to J.C. We declare we have no competing interests.

Author contributions

H.S. and I.J.L. conceived of and coordinated the study; J.C. carried out statistics and phylogenetic analyses; J.C., O.H., J.P. and J.M. generated genome size data; H.L., M.C., Y.R., S.Z. and M.G. provided critical material and chromosome counts; H.S. drafted the manuscript with the contribution of all authors. All authors gave final approval for publication.

References

Bainard J, Henry T, Bainard L, Newmaster S. 2011. DNA content variation in monilophytes and lycophytes: large genomes that are not endopolyploid. *Chromosome Research* **19**: 763–775.

Barker MS. 2013. Karyotype and genome evolution in pteridophytes. In: Greilhuber J, Doležel J, Wendel JF, eds. *Plant genome diversity, vol. 2*. Vienna, Austria: Springer, 245–253.

Barker MS, Wolf PG. 2010. Unfurling fern biology in the genomics age. *BioScience* **60**: 177–185.

Bellefroid E, Rambe SK, Leroux O, Viane RLL. 2010. The base number of 'loxoscaphoid' *Asplenium* species and its implication for cytoevolution in Aspleniaceae. *Annals of Botany* **106**: 157–171.

Bennett MD, Johnston S, Hodnett GL, Price HJ. 2000. *Allium cepa* L. cultivars from four continents compared by flow cytometry show nuclear DNA constancy. *Annals of Botany* **85**: 351–357.

Bennett MD, Leitch IJ. 2012. *Pteridophyte DNA C-values database* (release 5.0, Dec. 2012). [WWW document] URL http://data.kew.org/cvalues/. [accessed 1 August 2015]

Bomfleur B, McLoughlin S, Vajda V. 2014. Fossilized nuclei and chromosomes reveal 180 million years of genomic stasis in Royal Ferns. *Science* **343**: 1376–1377.

Bou Dagher-Kharrat M, Abdel-Samad N, Douaihy B, Bourge M, Fridlender A, Siljak-Yakovlev S, Brown SC. 2013. Nuclear DNA C-values for biodiversity screening: case of the Lebanese flora. *Plant Biosystems* **147**: 1228–1237.

Chang Y, Li J, Lu S, Schneider H. 2013. Species diversity and reticulate evolution in the *Asplenium normale* complex (Aspleniaceae) in China and adjacent areas. *Taxon* **62**: 673–687.

Christenhusz MJM, Chase MW. 2014. Trends and concepts in fern classification. *Annals of Botany* **113**: 571–594.

Christenhusz MJM, Zhang X-C, Schneider H. 2011. A linear sequence of extant families and genera of lycophytes and ferns. *Phytotaxa* **19**: 7–54.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.

Doležel J, Binarova P, Lucretti S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* **31**: 113–120.

Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, Obermayer R. 1998. Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Annals of Botany* **82**: 17–26.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2**: 2233–2244.

Doležel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* **85**: 625–631.

Duncan T, Smith AR. 1978. Primary basic chromosome numbers in ferns: facts or fantasies? *Systematic Botany* **3**: 105–114.

Dyer R, Pellicer J, Savolainen V, Leitch I, Schneider H. 2013. Genome size expansion and the relationship between nuclear DNA content and spore size in the *Asplenium monanthes* fern complex (Aspleniaceae). *BMC Plant Biology* **13**: 219.

Ebihara A, Ishikawa H, Matsumoto S, Lin S-J, Iwatsuki K, Takamiya M, Watano Y, Ito M. 2005. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. *American Journal of Botany* **92**: 1535–1547.

Ekrt L, Trávníček P, Jarolímová V, Vít P, Urfus T. 2009. Genome size and morphology of the *Dryopteris affinis* group in Central Europe. *Preslia* **81**: 261–280.

Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* **125**: 1–15.

Garland T, Dickerman AW, Janis CM, Jones JA. 1993. Phylogenetic analysis of covariance by computer-simulation. *Systematic Biology* **42**: 265–292.

Greilhuber J, Doležel J, Lysak MA, Bennett MD. 2005. The origin, evolution and proposed stabilization of the terms 'Genome size' and 'C-value' to describe nuclear DNA contents. *Annals of Botany* **95**: 255–260.

Harrell FE Jr. 2014. *Hmisc: Harell Miscellaneous.* R package version 3.14-4. [WWW document] URL http://CRAN.R-project.org/package=Hmisc. [accessed 1 August 2015].

Haufler CH. 1987. Electrophoresis is modifying our concepts of evolution in homosporous pteridophytes. *American Journal of Botany* **74**: 953–966.

Haufler CH. 2002. Homospory 2002: an odyssey of progress in pteridophyte genetics and evolutionary biology. *BioScience* **52**: 1081–1093.

Haufler CH. 2014. Ever since Klekowski: testing a set of radical hypotheses revives the genetics of ferns and lycophytes. *American Journal of Botany* **101**: 2036–2042.

Hennequin S, Ebihara A, Dubuisson J-Y, Schneider H. 2010. Chromosome number evolution in *Hymenophyllum* (Hymenophyllaceae), with special reference to the subgenus *Hymenophyllum*. *Molecular Phylogenetics and Evolution* **55**: 47–59.

Henry TA, Bainard JD, Newmaster SG. 2015. Genome size evolution in Ontario ferns (Polypodiidae): evolutionary correlations with cell size, spore size, and habitat type and an absence of genome downsizing. *Genome* **57**: 555–566.

Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY, Soltis PS *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–102.

Khandelwal S. 1990. Chromosome evolution in the genus *Ophioglossum* L. *Botanical Journal of the Linnean Society* **102**: 205–217.

Kim HT, Chung MG, Kim K-J. 2014. Chloroplast genome evolution in early diverged leptosporangiate ferns. *Molecules and Cells* **37**: 372–382.

Klekowski EJ, Baker HG. 1966. Evolutionary significance of polyploidy in the Pteridophyta. *Science* **153**: 305–307.

Lehtonen S. 2011. Towards resolving the complete fern tree of life. *PLoS ONE* **6**: e24851.

Leitch AR, Leitch IJ. 2012. Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist* **194**: 629–646.

Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651–663.

Leitch IJ, Leitch AR. 2013. Genome size diversity and evolution in land plants. In: Leitch IJ, Greilhuber J, Doležel J, Wendel JF, eds. *Plant genome diversity, vol. 2, physical structure, behaviour and evolution of plant genomes.* Wien, Austria: Springer-Verlag, 307–322.

Loureiro J, Rodriguez E, Doležel J, Santos C. 2007. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany* **100**: 87–888.

Lovis JD. 1977. Evolutionary patterns and processes in ferns. *Advances in Botanical Research* **4**: 229–415.

Maddison WP, Maddison DR. 2014. *Mesquite: a modular system for evolutionary analysis. v.2.75.* [WWW document] URL http://mesquiteproject.org [accessed 1 August 2015].

Manton I. 1950. *Problems of cytology and evolution in the Pteridophyta.* Cambridge, UK: Cambridge University Press.

Manton I, Vida G. 1968. Cytology of the fern flora of Tristan da Cunha. *Proceedings of the Royal Society of London. Series B* **170**: 361–379.

Mayrose I, Barker MS, Otto SP. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology* **59**: 132–144.

Midford PE, Garland JT, Maddison WP. 2011. *PDAP:PDTREE package 1.16.* [WWW document] URL http://mesquiteproject.org/pdap_mesquite/index.html [accessed 1 August 2015].

Nakazato T, Barker MS, Rieseberg LH, Gastony GL. 2008. Evolution of the nuclear genome of ferns and lycophytes. In: Ranker TA, Haufler CH, eds. *Biology and evolution of ferns and lycophytes.* Cambridge, UK: Cambridge University Press, 175–198.

Nitta JH, Ebihara A, Ito M. 2011. Reticulate evolution in the *Crepidomanes minutum* species complex (Hymenophyllaceae). *American Journal of Botany* **98**: 1782–1800.

Obermayer R, Leitch IJ, Hanson L, Bennett MD. 2002. Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Annals of Botany* **90**: 209–217.

Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* **53**: 673–684.

Pellicer J, Leitch IJ. 2014. The application of flow cytometry for estimating genome size and ploidy level in plants. In: Besse P, ed. *Molecular plant taxonomy: methods and protocols. Volume 1115.* Springer Science+Business Media: New York, NY, USA, 279–307.

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618–622.

Pryer KM, Schuettpelz E, Wolf PG, Schneider H, Smith AR, Cranfill R. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *American Journal of Botany* **91**: 1582–1598.

Pustahija F, Brown SC, Bogunic F, Basic N, Muratovic E, Ollier S, Hidalgo O, Bourge M, Stevanovic V, Siljak-Yakovlev S. 2013. Small genomes dominate in plants growing on serpentine soils in West Balkans, an exhaustive study of 8 habitats covering 308 taxa. *Plant and Soil* **373**: 427–453.

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. *Tracer v.1.6.* [WWW document] URL http://beast.bio.ed.ac.uk/Tracer [accessed 1 August 2015].

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient

Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Schneider H, Liu H, Clark J, Hidalgo O, Pellicer J, Zhang S, Kelly LJ, Fay MF, Leitch IJ. 2015. Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in the royal ferns. *New Phytologist* **207**: 10–13.

Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallon S, Lupia R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* **428**: 553–557.

Smith AR, Pryer KM, Schuettpelz E, Korall P, Schneider H, Wolf PG. 2006. A classification for extant ferns. *Taxon* **55**: 705–731.

Soltis DE, Albert VA, Leebes-Mack J, Bell CD, Paterson AH, Zheng CF, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidiy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.

Soltis DE, Soltis PS. 1987. Polyploidy and breeding systems in homosporous Pteridophyta: a reevaluation. *American Naturalist* **130**: 219–232.

Vanneste K, Sterck L, Myburg AA, Van de Peer Y, Mizrachi E. 2015. Horsetails are ancient polyploids: evidence from *Equisetum giganteum*. *Plant Cell* **27**: 1567–1578.

Wagner WH, Wagner FS. 1980. Polyploidy in pteridophytes. In: Lewis WH, ed. *Polyploidy: biological relevance*. New York, NY, USA: Plenum Press, 199–214.

Wang L, X-p Qi, Xiang Q-P, Heinrichs J, Schneider H, Zhang X-C. 2010. Phylogeny of the paleotropical fern genus *Lepisorus* (Polypodiaceae, Polypodiopsida) inferred from four chloroplast DNA regions. *Molecular Phylogenetics and Evolution* **54**: 211–225.

Webb CO, Ackerly DD, Kembel SW. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24**: 2098–2100.

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA *et al.* 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* **111**: E4859–E4868.

Williams EW, Waller DM. 2012. Phylogenetic placement of species within the genus *Botrychium* s.s. (Ophioglossaceae) on the basis of plastid sequences, amplified fragment length polymorphisms, and flow cytometry. *International Journal of Plant Sciences* **173**: 516–531.

Wolf PG, Sessa EB, Marchant DB, Li F-W, Rothfels CJ, Sigel EM, Gitzendanner MA, Visger CJ, Banks JA, Soltis DE *et al.* 2015. An exploration into fern genome space. *Genome Biology and Evolution* **7**: 2533–2544.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* **106**: 13875–13879.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Summary of the results inferring the impact of phylogenetic uncertainty on the Bayesian estimation of genome size for the five basal orders of ferns (Equisetales, Marattiales, Ophioglossales, Osmundales and Psilotales).

Fig. S2 Phylogenetic hypothesis used to infer the evolution of genome size (1C), chromosome number (*2n*) and chromosome size (2C/*2n*).

Table S1 Summary of the information obtained and used for the present study

Table S2 Estimated parameter values using chromosome evolve

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Clark JW, Donoghue PCJ. 2017 Constraining the timing of whole genome duplication in plant evolutionary history. *Proc. R. Soc. B* **284**: 20170912. http://dx.doi.org/10.1098/rspb.2017.0912

Received: 27 April 2017
Accepted: 1 June 2017

Subject Category:
Palaeobiology

Subject Areas:
evolution, genomics, plant science

Keywords:
genome duplication, plant evolution, polyploidy, molecular clock

Authors for correspondence:
James W. Clark
e-mail: james.clark@bristol.ac.uk
Philip C. J. Donoghue
e-mail: phil.donoghue@bristol.ac.uk

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3807310.

THE ROYAL SOCIETY PUBLISHING

Constraining the timing of whole genome duplication in plant evolutionary history

James W. Clark and Philip C. J. Donoghue

School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

JWC, 0000-0003-2896-1631; PCJD, 0000-0003-3116-7463

Whole genome duplication (WGD) has occurred in many lineages within the tree of life and is invariably invoked as causal to evolutionary innovation, increased diversity, and extinction resistance. Testing such hypotheses is problematic, not least since the timing of WGD events has proven hard to constrain. Here we show that WGD events can be dated through molecular clock analysis of concatenated gene families, calibrated using fossil evidence for the ages of species divergences that bracket WGD events. We apply this approach to dating the two major genome duplication events shared by all seed plants (ζ) and flowering plants (ε), estimating the seed plant WGD event at 399–381 Ma, and the angiosperm WGD event at 319–297 Ma. These events thus took place early in the stem of both lineages, precluding hypotheses of WGD conferring extinction resistance, driving dramatic increases in innovation and diversity, but corroborating and qualifying the more permissive hypothesis of a 'lag-time' in realizing the effects of WGD in plant evolution.

1. Background

The discovery in plant genomes of evidence of recurrent whole genome duplication events (WGD; polyploidy) has reignited debate over its importance in land plant evolution [1,2]. Several causal hypotheses have emerged linking WGD to key innovations [3], increased rates of diversification [4] and extinction resistance that may have facilitated the success of multiple lineages of extant plants [5]. The mechanisms through which genome duplication can result in evolutionary novelty are becoming better understood and the traditional models of neo- and subfunctionalization have now been hybridized with models of dosage balance in attempts to explain how evolutionary innovation can arise post-WGD in the face of extensive gene loss and stabilizing patterns of gene retention [6,7]. Furthermore, there now exist elegant examples of genes and gene families that have taken on new functions (neofunctionalization) following multiple rounds of WGD and then playing a key role in the evolution of plant lineages [8]. The link between polyploidy and diversification remains controversial [9], but there exists some evidence that several of the ancient WGD events in angiosperms correlate with shifts in diversification [4]. Separating the WGD events and the shifts in diversification are a 'lag' of several million years, which has been explained as the period of fractionation post-WGD and, in turn, the feature of WGD that leads to innovation and diversification [10]. However, at the broadest scale, these hypotheses are underpinned by the relative phylogenetic placement and absolute timing of each event. Though the relative phylogenetic timing of plant WGD events is well constrained, their absolute timing is not [9].

Constraining the phylogenetic position of WGD events relies on broad taxonomic sampling of genomic or transcriptomic data. The presence or absence of shared 'age peaks' in Ks plots of synonymous substitution rates between duplicates provides evidence for shared genome duplications [11]. This approach culminated in a survey of 41 plant genomes focusing on angiosperms [5] and more recently several transcriptomes also highlighting the presence of WGD within the evolutionary history of gymnosperms [12] and peat mosses [13].

The number and position of the peaks on the Ks plot also reveals the relative timing of each event, with multiple peaks representing multiple successive WGDs. The absolute timing of each event can be obtained indirectly by phylogenetically bracketing the event—the event must have occurred along the branch between those lineages that have undergone the WGD and those that have not. However, despite well-sampled exceptions among certain groups of angiosperms [14–16], there are few cases where the sampling of taxa is dense enough to prevent very long branches, and so the ages of genome duplication events must be inferred directly. Direct dates can be obtained by converting the relative timing of peaks on a Ks plot into absolute ages. This has the advantage that it does not require additional taxon sampling and so estimates can be obtained for WGD events isolated on long branches [17]. A major caveat of this approach is that it relies on the assumption of a strict molecular clock that, depending on shifts in the rate of sequence evolution, can lead to inaccurate age estimates. Furthermore, Ks plots are known to saturate beyond a certain age, meaning that they cannot always distinguish more ancient duplications and may lead to artificial peaks in the distribution [18]. More complex relaxed clock methods can be employed in a phylogenetic or phylogenomic approach, whereby the individual gene families containing signal of WGD are reconstructed and individually dated [19]. The distribution of ages obtained can then be plotted to provide a range of estimates for each event. This approach is more powerful and has been used to estimate the ages of multiple WGD events across the angiosperms, where genomic and transcriptomic data are more abundant [19,20]. However, dating individual gene trees does not fully exploit the power of the molecular clock and the power of individual gene trees is likely to diminish over longer periods of evolutionary time. Increasing the amount of sequence data by concatenating multiple gene families into alignments decreases uncertainty in the estimation of relative ages [21], and can be used to date the absolute timing of WGD events [22] yet, to date, studies focusing on WGD in plants have relied on the power of individual gene trees. Directly dating WGD events using concatenated gene trees also provides estimates of the absolute timing of the WGD in relation to subsequent speciation events within the lineage, since gene trees observe species divergences as well as duplication events. Thus, concatenated gene trees have the potential to provide an accurate estimate of the absolute timing of WGD events relative to the diversification events in which they are causally implicated.

The seed plants (Spermatophyta) are the most species rich of extant plant clades, encompassing the gymnosperms and angiosperms (flowering plants). WGD events have been identified at the base of all seed plants (ζ; [12,20]) and at the base of all angiosperms (ε; [20]), and so all extant flowering plants have undergone at least two rounds of genome duplication. Previous attempts to date these events were based on distributions of ages inferred using poorly defined calibrations and penalized likelihood molecular clock methods [20] that have since been found unreliable [23]. The WGD shared by all extant angiosperms has been linked with the 'big bang' diversification of the Mesangiospermae (following a lag period) as well as several major innovations, including the origin of the flower [3,4]. WGD has been thought to be less prevalent within gymnosperms, the sister clade to angiosperms (together comprising Spermatophyta), despite the fact that the ζ WGD is part of their shared evolutionary history. More recent evidence has indicated that WGD has occurred in several gymnosperm lineages and confirmed that the ζ WGD (spermatophyte) was not shared with their sister lineage, the ferns [12].

Conventionally molecular clock dating approaches have sought to minimize the influence of duplication by using only single copy genes. In contrast, we exploit the pattern of paralogy produced by WGD in the evolutionary history of multiple gene families and concatenate them into a partitioned alignment. Combined with broad taxon sampling and multiple fossil calibrations, we demonstrate an approach for dating gene trees to provide well-constrained estimates of the timing of duplication events and attendant speciation events.

2. Material and methods

Gene families containing signal of the ζ (spermatophyte) and ε (angiosperm) WGD events and those that contain the signal of both were catalogued by Jiao *et al*. [4], and from these we expanded orthogroups by obtaining amino acid sequences using Plaza 3.0 (bioinformatics.psb.ugent.be/plaza), and GreenPhyl 4 (www.greenphyl.org). Further sequences were obtained by local BLAST searches of iPlant (www.iplantcollaborative.org). One hundred and twenty-eight species were sampled in total, representing all major lineages of land plants and these are listed in electronic supplementary material, table S1. Four datasets were assembled for all taxa: families containing a clear signal of just the ε WGD event (angiosperm dataset), just the ζ WGD event (spermatophyte dataset), families containing signal of both events ($\zeta + \varepsilon$ dataset), and a combined dataset. To verify a clear signal of the relevant WGD event in each gene family, we built individual gene trees based on multiple amino acid sequence alignments generated using MAFFT while model selection and gene tree reconstructions were performed using IQ-TREE [24]. We opted for a conservative approach, discarding orthogroups that following phylogenetic reconstruction and visual inspection did not clearly reflect the signal of either or both WGDs (e.g. electronic supplementary material, figure S1), had sequence alignments shorter than 100 amino acids, displayed a topology that was incongruent with our current understanding of land plant phylogeny with either the total group seed plants or major lineages within being resolved as non-monophyletic, or were too large with multiple nested duplications, resulting in large numbers of sequences having to be discarded. Of 130 orthogroups surveyed, 12 gene families were found containing a clear signal of the ε WGD. The number of sequences among individual gene families ranged from 87–126 and when concatenated a total of 176 tips. Fourteen further gene families were found for the ζ WGD, representing 189 tips when concatenated and varying from 106 to 149 tips individually. An additional seven gene families were found containing the signal for both, for which 254 tip sequences were assembled when concatenated and individual gene families ranging from 132 to 249 tips. The combined dataset contained 33 gene families, with one node representing ζ, but two representing ε. As 12 gene families contain only one node with the ε duplication, the event was represented only once in the combined analysis, to maximize precision at this node. Similarly, angiosperm gene copies from gene families not containing signal of the ε duplication were randomly assigned to one side of the duplication. Due to differential retention, a copy of each gene paralogue was not present in all families and the number of tips in each gene family is listed in electronic supplementary material, table S3.

Across all analyses, nodes were constrained using 35 fossil calibrations spanning land plant phylogeny defined using best practice [25] (electronic supplementary material, table S2). The

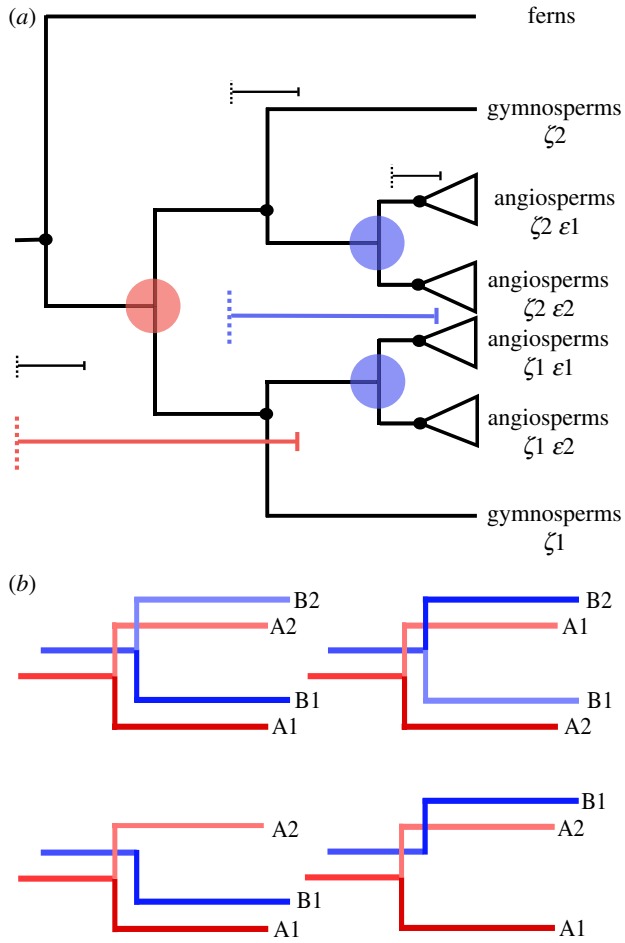


Figure 1. (*a*) An example gene tree showing the seed plant (ζ, red) and angiosperm (ε, blue) duplications. The duplication events are constrained using minima and maxima (coloured brackets) based on fossils used to constrain speciation events (black brackets). (*b*) Gene trees may retain both copies of the duplicate gene (top), or a single copy may be lost (bottom). When concatenating duplicates from different gene families, given that both copies are descended from the same event, their assignment to either side of the duplication is arbitrary. (Online version in colour.)

duplication nodes were constrained temporally to reflect the possibility of the WGD occurring at any point following the divergence of spermatophytes from an ancestral euphyllophyte (ζ WGD event) and for angiosperms from an ancestral spermatophyte (ε WGD event) (figure 1). Calibrations that provided only a minimum age were modelled as a hard minimum bound with a truncated Cauchy distribution ($p = 0.1, c = 0.2$). Calibrations that provided a maximum age were modelled with a soft maximum with a uniform distribution between the minimum and maximum age [26]. Molecular clock analyses were conducted on concatenated alignments using the normal approximation method in MCMCtree under the appropriate model [27]. The normal approximation method provides a fast and efficient way of analysing large datasets using complex models and a relaxed clock and is run under a fixed topology. We ran all analyses on a topology reflecting both WGD events and recent hypotheses of relationships among land plants [28] (electronic supplementary material, figure S2). We also reconstructed the topology based on our own datasets using IQ-TREE and found that it was highly congruent with the constraint tree. Each analysis was run twice independently and regularly checked for convergence and for effective sample sizes greater than 200 using Tracer v. 16 [29].

Assuming autopolyploidy, each WGD event produces two daughter nodes that are created simultaneously and that must have the same age, and so the assignment of each paralogue to either node of the duplication is arbitrary (figure 1). In this way paralogues between the gene families can be concatenated in multiple combinations, so long as they are consistent within each gene family. To explore the impact of different combinations of paralogy groups between gene families, we randomly reassigned groups to either node using the $\zeta + \varepsilon$ dataset containing both duplications.

The extent to which the low number of available gene families impacted on the estimation of dates was explored through infinite sites analyses [30]. The gene families were successively concatenated and the analysis repeated with one more gene family each time. The relationship between the mean age estimates and the widths of the 95% HPDs was used as a measure of the precision of the data versus the uncertainties induced by the fossil calibrations. Higher R^2 values indicate that large HPD widths are due to increasing uncertainty in the fossil record deeper in time. A saturation of the curve suggests that adding further sequence data would not increase the precision of the analysis, since it is limited by the information available in the fossil record.

3. Results

In most Bayesian molecular software, specified node age priors are modified in the construction of the joint time prior to achieve the expectation that only ages compatible with the assumption that ancestral nodes are older than their descendants, are proposed to the MCMC [31,32]. To ensure that these effective priors are biologically reasonable, we estimated them by running the analysis without sequence data. The effective priors are compatible with the original palaeontological and phylogenetic evidence, yielding broad 95% HPDs for the timing of WGDs in all analyses, though both were truncated relative to the specified calibrations. The spans of the 95% HPD for the prior on the ζ and ε WGD events are 81 (434–353 Ma) and 111 (355–244 Ma) million years, respectively (table 1). In the separate analyses of both the ζ and the ε WGD events, the truncation effects on the prior were the same as for the combined analysis, and so the additional nodes in the combined analysis and the $\zeta + \varepsilon$ dataset did not affect the effective prior.

In all instances, the addition of sequence data yielded estimates congruent with, yet more precise than, the joint time prior. Estimates for both WGD events were compared between gene families using the $\zeta + \varepsilon$ dataset, and we found variation in both the width of the 95% HPD and the absolute age estimates, though the overlapping distributions of the HPDs showed that the gene families were congruent. While some gene families produced much more precise estimates, the variation in estimates between all gene families showed a similar level of precision to the joint time prior alone, ranging from 435–346 Ma for the ζ WGD event and 355–244 for the ε WGD event. The $\zeta + \varepsilon$ dataset also allowed us to compare the estimates for the ε duplication, which is represented twice in each gene family, within gene families. We found that the 95% HPD widths for the event varied within gene families, though this is likely due to the absence of paralogues on one side of the duplication. The only family with all paralogues present, CDK, showed estimates consistent in both age and uncertainty across both nodes.

The greatest effect in terms of precision was produced by increasing the amount of sequence data by concatenating the gene families. The effect of missing paralogues across both duplication nodes in the $\zeta + \varepsilon$ dataset was minimized and

Table 1. Ninety-five per cent HPD estimates for the age of both WGD events, summarizing the effective prior, individual gene families (1 to 7), the effects of concatenating gene families, the expanded and combined datasets.

node	effective prior	gene families							concatenated gene families						ε dataset	ζ dataset	combined dataset
		1	2	3	4	5	6	7	1–2	1–3	1–4	1–5	1–6	1–7			
spermatophyte duplication (ζ)	353–434	382–435	346–411	346–411	354–418	354–404	357–415	355–433	390–433	386–430	380–418	380–416	377–408	378–409	—	380–401	381–399
angiosperm duplication (ε) ζ1 ε	244–355	270–339	250–353	248–328	280–354	258–340	249–351	254–356	273–336	268–323	280–323	285–331	282–325	281–323	295–321	—	297–319
angiosperm duplication (ε) ζ2 ε	244–355	267–340	273–344	247–350	245–349	277–362	247–313	245–355	278–333	276–330	276–322	289–338	283–325	276–321	—	—	—

the age estimates for both ε nodes were highly consistent. The ζ + ε concatenation was also considerably more precise than any of the individual gene families (table 1). Multiple concatenations were tested on this dataset, to determine if the assignment of paralogues between duplicates affected the estimates. We did not observe any material differences in age or uncertainty, indicating that the results are robust to the way in which the gene families are concatenated.

The addition of further sequence data for each duplication event in turn produced results of even greater precision. The angiosperm dataset estimated an age of 321–295 Ma for the ε WGD event, almost five times more precise than the joint time prior alone. A similar increase in precision was obtained by the spermatophyte dataset, the ζ duplication estimated to have occurred 400–380 Ma, four times more precise than the joint time prior alone. Based on the largest amount of data, the combined analysis of the combined dataset produced results that were highly congruent with the two individual datasets, if not marginally more precise, estimating 399–381 Ma and 319–297 Ma for the ζ and ε WGD events, respectively (figure 2).

Infinite sites plots suggest that though the R^2 value showed little changed with increased sequence data, the addition of sequence data reduced the uncertainty of estimates (figure 3). With 19 gene families, the amount of error was continuing to decrease, suggesting that additional gene families may increase precision further.

4. Discussion

(a) Inferring the age of whole genome duplication

Our results indicate that the evolutionary history of gene families can be exploited to obtain precise estimates of the age of WGD events. These methods depend on both careful selection of fossil constraints and available gene families containing signal of WGD events, though even with limited sequence data, we greatly improve the precision over the raw calibrations alone.

Both the ε (angiosperm) and ζ (spermatophyte) genome duplication events have been independently reported [12,20], yet we were unable to find large numbers of gene families with clear signal of either or both events. The paucity of available gene families for these WGD events is likely in part a result of our conservative criteria in selecting gene families based on topology. In part, this reflects the limitations of single genes to resolve unequivocal phylogenetic signal for such events over long timescales. However, it also reflects the antiquity of the events, given that retention of genes following a WGD follows a decay pattern and widespread gene loss leads to a gradually decreasing phylogenetic signal over time. It is unsurprising that so few gene families remain with a clear signal of these events and, when considered next to existing evidence for these events [12,20], our findings are entirely compatible with the ε and ζ duplication events. Our results indicate that the evolutionary history of gene families can be exploited to obtain precise estimates of the age of WGD events. Infinite sites plots lead us to expect that the addition of further sequence data will leverage further precision. Similarly, WGD events that are more recent and may contain more genome-wide data, may be dated using the same approach but with greater precision.

Unlike genomic datasets that can be used for gene-tree reconciliation and the construction of Ks plots, the methods

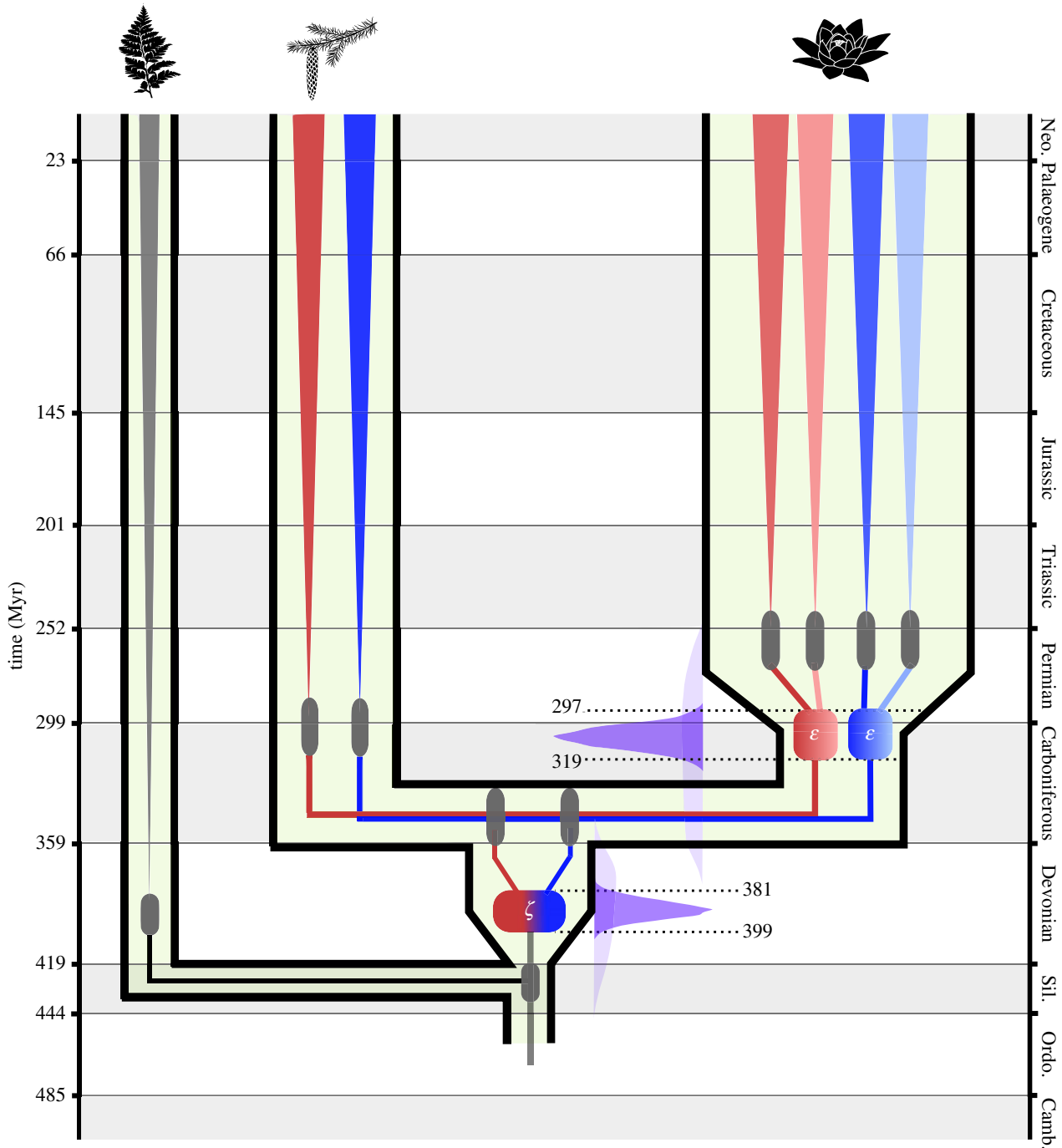


Figure 2. Estimated dates for the occurrence of both the seed plant (ζ) and angiosperm (ε) duplication events based on a molecular clock analysis of 33 concatenated gene families. Age estimates (95% HPD) for the divergences of the major lineages and crown groups represented by grey bars. The age estimates (95% HPD) of two duplication events are represented by coloured boxes, with the subsequent subgenomes represented first by blue and red (ζ), then by lighter and darker shades of each colour (ε). For each duplication event, the effective prior is shown (light blue) next to the posterior distribution (dark blue). (Online version in colour.)

presented here focus solely on the dating of WGD events, rather than their characterization. However, the congruence of age estimates between gene families serves as a test of their coincidence, as anticipated by WGD. The annotation of gene families to either side of the duplication event requires greater care and is a potentially limiting factor on the number of gene families that can be analysed, yet we have demonstrated that even with a relatively small dataset (compared to a genomic dataset), high levels of precision can be achieved. Novel molecular clock approaches such as cross bracing could also be used to increase precision around the duplication nodes, especially as they are so difficult to constrain [33].

An additional caveat is that WGD or polyploidy is often categorized into two distinct classes [34], autopolyploidy and allopolyploidy, traditionally distinguished based on the number of parent species, but also characterized by the patterns of fractionation post-WGD. The mode of duplication may impact our estimates of duplication age [35], as the point at which duplicates coalesce is actually the timing of divergence of the two parental species, or a more ancestral autoploidy event, as opposed to the alloploidy event itself [35]. New methods are emerging to discriminate between auto- and allopolyploidy [36], but these are likely to fail when applied to more ancient genome duplication events. However, allopolyploidy

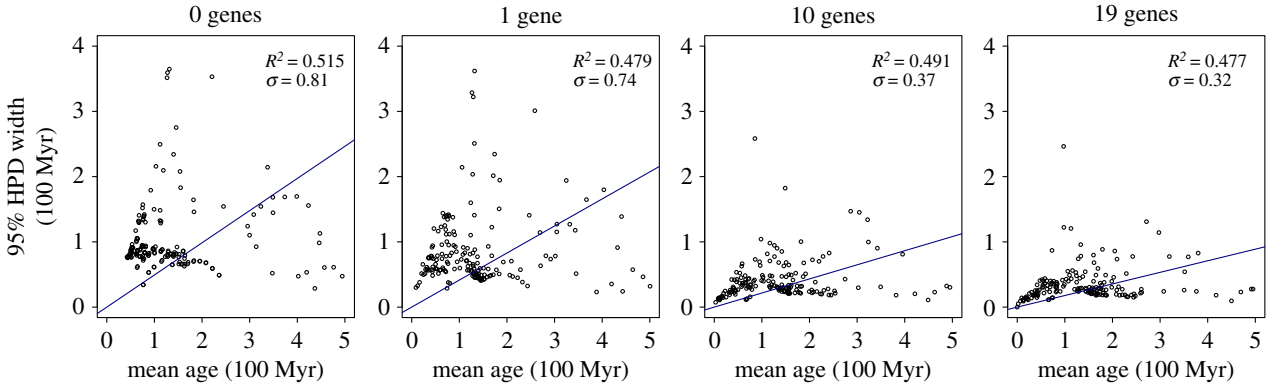


Figure 3. Infinite sites plots for the most complete (angiosperm) dataset, with the regression between the mean age and the 95% HPD shown for 0, 1, 10 and 19 gene datasets. The R^2 and error terms are also shown. (Online version in colour.)

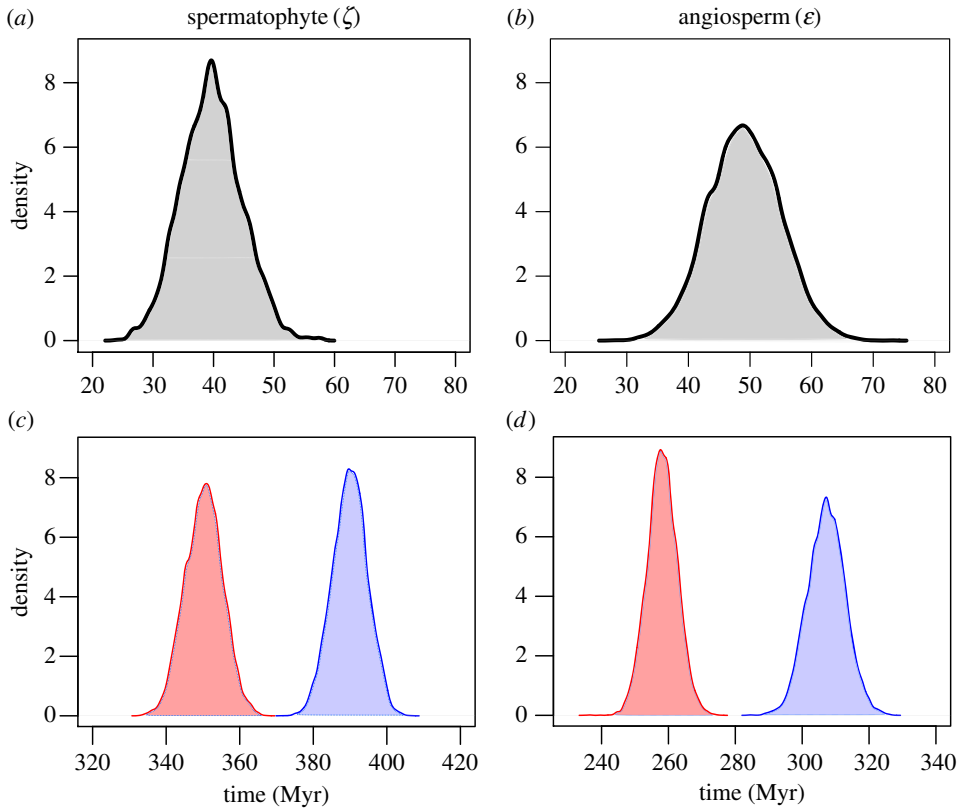


Figure 4. The posterior probabilities of (*a*) the lag between the ζ duplication and the diversification of crown spermatophytes and (*b*) the lag between the ε duplication and the diversification of crown angiosperms. The posterior probabilities of the absolute age of the WGD events (blue) and diversification (red) are also shown for (*c*) ζ and spermatophytes and (*d*) ε and angiosperms. (Online version in colour.)

would only have a large impact on accuracy if hybridization occurred between very distant parent species.

(b) Dating duplication, diversification and innovation

Our most comprehensive analysis of 33 gene families indicated that the genome duplication present in all crown spermatophytes occurred 399–381 Ma, a period spanning the Early to Late Devonian (figure 2). The WGD event present in all crown angiosperms occurred almost 100 Myr later, 319–297 Ma, across the Carboniferous–Permian boundary (figure 2). Gene trees contain both the signal of WGD and species divergence, allow a direct estimation of the age of the WGD event relative to the age of the crown group (figure 4). Both estimates predict that the respective WGD events occurred early in the stem of both lineages, predating the diversification of the crown group by about 50 Myr. These estimates are considerably older than those of Jiao *et al.* [20], yet our estimates for the age of the seed plant (360–340 Ma) and angiosperm (267–247) crown groups are comparable to other molecular clock analyses [37,38], allowing us to reject the notion that the duplications occurred late in the stem lineage. Greater precision in the absolute age of WGD events leveraged by concatenation allows that hypotheses can be more rigorously tested. WGD occurring early in the stem lineage has two implications for current hypotheses regarding the role of WGD in plant evolution.

First is the hypothesis that WGD drives evolutionary success [39–41], or confers extinction resistance [19,42], since the long stem lineages of both groups are, by definition, characterized by extinction. However, many extinct lineages must also share these genome duplications. For example, the ζ

duplication predates the appearance of the earliest seed plants, the pteridosperms and cordaitales, and so WGD cannot have contributed to their diversification or conferred extinction resistance, as has been proposed for the ancient palaeopolyploid *Equisetum* [17]. The long-term evolutionary success of seed plants and especially angiosperms is unquestionable, and there is considerable evidence for the role of gene duplication in the evolution of angiosperms, in particular [3,43], yet our results are more in keeping with the idea of 'rarely successful polyploids' [39]. The challenges faced by polyploids in order to establish and persist may be partially responsible for extinctions in a lineage post-WGD, and it may be the case that extant spermatophytes and angiosperms are the surviving lineages best able to exploit any long-term competitive advantages [42]. Secondly, if their crown clades of seed and flowering plants can be considered to be characterized by evolutionary success, this has been achieved in both lineages after a substantial lag post-WGD. Our results indicate that the lag between the ζ WGD event and the divergence of crown spermatophytes is 22–60 Myr, and 27–65 Myr between the ε WGD event and the divergence of crown angiosperms (figure 4). These are comparable to the results of Tank *et al*. [4], who estimated a 49.2 Myr lag between the ε WGD event and the shift in diversification of angiosperms, though without directly inferring the age of the WGD. Tank *et al*. [4] also estimated that the rate shift in diversification among angiosperms occurred at 213 Ma, following the divergence of Mesangiospermae which, following our age estimates, indicates a lag of 84–106 Myr. Ultimately, these results indicate that more precise age estimates require more precise hypotheses regarding the role of WGD in promoting evolutionary success. Given these long lag periods and that some, though clearly not all, clades that share a history of WGD are diverse or characterized by innovations, it requires more explicit hypotheses regarding which clades are considered successful.

Evidently, we find no direct support for the deterministic role of WGD in driving diversification or innovation. Rather, our data are more compatible with the more permissive model of evolution via genome duplication that emphasizes the importance of the post-WGD period of genome fractionation. During this period, the need to maintain a dosage balance of protein products selects for the maintenance of duplicates, followed by a relaxation of selection allowing sub- and neofunctionalization [7]. An additional consideration is the lineage specific re-diploidization model, which applies when species divergence occurs before the diploidization process in complete [44]. Under this model, the lag is produced by the pattern of tetrasomic inheritance that is characteristic of autopolyploidy, leading to massively delayed functional divergence of duplicate genes. This model also predicts that duplicate genes evolve independently in separate lineages, and that this can explain the divergent evolutionary trajectories of lineages that share the same history of WGD [44]. This more permissive model explains the 'long fuse' or 'lag' found in our results, whereby an early WGD during a lineage's evolution provides a primer for subsequent innovation and diversification, leading to the evolutionary success of both lineages [42]. It also explains the paucity of genes preserving all paralogues anticipated as a phylogenetic footprint of the ζ and ε WGD events, as a consequence of post-duplication dysploidy leading to dosage bias.

The quantification of this lag is clearly relevant to understanding the role of WGD in plant evolution [42]. Our methods are applicable to other WGD events characterized previously within the plant kingdom, including those thought to be associated with increased diversification or the K–Pg boundary [4,5]. Furthermore, these methods could be used to clarify the timing of the proposed WGDs associated with the origins and early evolution of vertebrates [45], which are still undermined by uncertainty around their timing.

5. Conclusion

Accurate and precise estimates of the timing of WGD events are fundamental to our understanding their significance on a macroevolutionary scale and can be achieved by coupling a careful appraisal of the fossil record with molecular clock approaches. We demonstrated that by concatenating multiple gene families with a shared history of WGD into a single alignment, the ages of two ancient WGD events, ε (angiosperm) and ζ (spermatophyte), were estimated to a high degree of precision. Both events were found to occur early in the stem of each lineage, predating the divergence of the crown groups by 50 Myr. These methods can be applied to date any previously characterized WGD event, including those identified in yeasts and vertebrates.

Data accessibility. Electronic supplementary material includes Supplemental Experimental Procedures, three figures and three tables and can be found with this article online. The molecular sequence alignment and trees with fossil calibrations have been deposited in Figshare: https://figshare.com/s/d46377d5ae6999c0cd52.

Authors' contributions. J.W.C. and P.C.J.D. conceived the project and designed the analysis. J.W.C. prepared the datasets and performed the analyses. J.W.C. and P.C.J.D. interpreted the results. J.W.C. wrote the main draft of the manuscript, to which P.C.J.D. contributed.

Competing interests. The authors declare no competing interests.

Funding. This research was funded by Biotechnology and Biosciences Research Council (UK) grant nos. (BB/N000609/1; BB/N000919/1), Natural Environment Research Council grant no. (NE/N002067/1), the Royal Society, and the Wolfson Foundation.

Acknowledgements. We thank members of the Bristol Palaeobiology Research Group for discussion and two anonymous reviewers for their comments and improvements to the manuscript.

References

1. Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS, Rieseberg LH, Otto SP. 2015 Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis *et al*. (2014). *New Phytol.* **206**, 27–35. (doi:10.1111/nph.13192)
2. Soltis DE *et al*. 2014 Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al*. (2011). *New Phytol.* **202**, 1105–1117. (doi:10.1111/nph.12756)
3. Soltis PS, Soltis DE. 2016 Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–65. (doi:10.1016/j.pbi.2016.03.015)
4. Tank DC *et al*. 2015 Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* **207**, 454–467. (doi:10.1111/nph.13491)

5. Vanneste K, Baele G, Maere S, Van de Peer Y. 2014 Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* **24**, 1334–1347. (doi:10.1101/gr.168997.113)
6. Teufel AI, Liu L, Liberles DA. 2016 Models for gene duplication when dosage balance works as a transition state to subsequent neo- or sub-functionalization. *BMC Evol. Biol.* **16**, 45. (doi:10.1186/s12862-016-0616-1)
7. Conant GC, Birchler JA, Pires JC. 2014 Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* **19**, 91–98. (doi:10.1016/j.pbi.2014.05.008)
8. Edger PP *et al.* 2015 The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl Acad. Sci. USA* **112**, 8362–8366. (doi:10.1073/pnas.1503926112)
9. Kellogg EA. 2016 Has the connection between polyploidy and diversification actually been tested? *Curr. Opin. Plant Biol.* **30**, 25–32. (doi:10.1016/j.pbi.2016.01.002)
10. Dodsworth S, Chase MW, Leitch AR. 2016 Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* **180**, 1–5. (doi:10.1111/boj.12357)
11. Lynch M, Conery JS. 2000 The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155. (doi:10.1126/science.290.5494.1151)
12. Li Z *et al.* 2015 Early genome duplications in conifers and other seed plants. *Sci. Adv.* **1**, e1501084. (doi:10.1126/sciadv.1501084)
13. Devos N, Weston DJ, Rothfels CJ, Johnson MG, Shaw AJ. 2016 Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytol.* **211**, 300–318. (doi:10.1111/nph.13887)
14. Estep MC *et al.* 2014 Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl Acad. Sci. USA* **111**, 15 149–15 154. (doi:10.1073/pnas.1404177111)
15. Barker MS, Li Z, Kidder TI, Reardon CR, Lai Z, Oliveira LO, Scascitelli M, Rieseberg LH. 2016 Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am. J. Bot.* **103**, 1203–1211. (doi:10.3732/ajb.1600113)
16. Kagale S *et al.* 2014 Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* **26**, 2777–2791. (doi:10.1105/tpc.114.126391)
17. Vanneste K, Sterck L, Myburg AA, Van de Peer Y, Mizrachi E. 2015 Horsetails are ancient polyploids: evidence from *Equisetum giganteum*. *Plant Cell* **27**, 1567–1578. (doi:10.1105/tpc.15.00157)
18. Vanneste K, Van de Peer Y, Maere S. 2013 Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* **30**, 177–190. (doi:10.1093/molbev/mss214)
19. Fawcett JA, Maere S, Van de Peer Y. 2009 Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl Acad. Sci. USA* **106**, 5737–5742. (doi:10.1073/pnas.0900906106)
20. Jiao Y *et al.* 2011 Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100. (doi:10.1038/nature09916)
21. dos Reis M, Donoghue PCJ, Yang Z. 2016 Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80. (doi:10.1038/nrg.2015.8)
22. Macqueen DJ, Johnston IA. 2014 A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. R. Soc. B* **281**, 20132881. (doi:10.1098/rspb.2013.2881)
23. Thorne JL, Kishino H. 2005 Estimation of divergence times from molecular sequence data. In *Statistical methods in molecular evolution* (ed. R Nielsen), pp. 233–256. New York, NY: Springer.
24. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
25. Parham JF *et al.* 2012 Best practices for justifying fossil calibrations. *Syst. Biol.* **61**, 346–359. (doi:10.1093/sysbio/syr107)
26. Warnock RC, Parham JF, Joyce WG, Lyson TR, Donoghue PCJ. 2015 Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B* **282**, 20141013. (doi:10.1098/rspb.2014.1013)
27. Yang Z. 2007 PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
28. Wickett NJ *et al.* 2014 Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. USA* **111**, E4859–E4868. (doi:10.1073/pnas.1323926111)
29. Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014 Tracer v1.6. See http://beast.bio.ed.ac.uk/Tracer.
30. Yang Z, Rannala B. 2006 Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226. (doi:10.1093/molbev/msj024)
31. Inoue JG, Donoghue PCJ, Yang Z. 2010 The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**, 74–89. (doi:10.1093/sysbio/syp078)
32. Warnock RCM, Yang Z, Donoghue PCJ. 2012 Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* **8**, 156–159. (doi:10.1098/rsbl.2011.0710)
33. Shih PM, Matzke NJ. 2013 Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc. Natl Acad. Sci. USA* **110**, 12 355–12 360. (doi:10.1073/pnas.1305813110)
34. Garsmeur O, Schnable JC, Almeida A, Jourda C, Hont A, Freeling M. 2014 Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* **31**, 448–454. (doi:10.1093/molbev/mst230)
35. Doyle JJ, Egan AN. 2010 Dating the origins of polyploidy events. *New Phytol.* **186**, 73–85. (doi:10.1111/j.1469-8137.2009.03118.x)
36. Marcet-Houben M, Gabaldón T. 2015 Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* **13**, e1002220. (doi:10.1371/journal.pbio.1002220)
37. Murat F, Armero A, Pont C, Klopp C, Salse J. 2017 Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.* **49**, 490–496. (doi:10.1038/ng.3813)
38. Foster CSP *et al.* 2017 Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* **66**, 338–351.
39. Arrigo N, Barker MS. 2012 Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* **15**, 140–146. (doi:10.1016/j.pbi.2012.03.010)
40. Madlung A. 2013 Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**, 99–104. (doi:10.1038/hdy.2012.79)
41. Soltis DE, Visger CJ, Soltis PS. 2014 The polyploidy revolution then. . .and now: Stebbins revisited. *Am. J. Bot.* **101**, 1057–1078. (doi:10.3732/ajb.1400178)
42. Fawcett JA, Van de Peer Y. 2010 Angiosperm polyploids and their road to evolutionary success. *Trends Evol. Biol.* **2**, 3. (doi:10.4081/eb.2010.e3)
43. Chanderbali AS 2016 Evolving ideas on the origin and evolution of flowers: new perspectives in the genomic era. *Genetics* **202**, 1255–1265. (doi:10.1534/genetics.115.182964)
44. Robertson FM *et al.* 2017 Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* **18**, 111. (doi:10.1186/s13059-017-1241-z)
45. Donoghue PCJ, Purnell MA. 2005 Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.* **20**, 312–319. (doi:10.1016/j.tree.2005.04.008)

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Puttick MN *et al*. 2017 Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proc. R. Soc. B* **284**: 20162290.
http://dx.doi.org/10.1098/rspb.2016.2290

Received: 19 October 2016
Accepted: 5 December 2016

Subject Category:
Palaeobiology

Subject Areas:
evolution, palaeontology, taxonomy and systematics

Keywords:
phylogeny, Bayesian, parsimony, cladistics, morphology, palaeontology

Authors for correspondence:
Davide Pisani
e-mail: davide.pisani@bristol.ac.uk
Philip C. J. Donoghue
e-mail: phil.donoghue@bristol.ac.uk

†These authors contributed equally to this study.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3653186.

THE ROYAL SOCIETY PUBLISHING

Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data

Mark N. Puttick[1,3,†], Joseph E. O'Reilly[1,†], Alastair R. Tanner[2], James F. Fleming[1], James Clark[1], Lucy Holloway[1], Jesus Lozano-Fernandez[1,2], Luke A. Parry[1], James E. Tarver[1], Davide Pisani[1,2] and Philip C. J. Donoghue[1]

[1]School of Earth Sciences, and [2]School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK
[3]Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK

MNP, 0000-0002-1011-3442; JEO, 0000-0001-9775-253X; JC, 0000-0003-2896-1631; LH, 0000-0003-1603-2296; JL-F, 0000-0003-3597-1221; LAP, 0000-0002-3910-0346; PCJD, 0000-0003-3116-7463

Morphological data provide the only means of classifying the majority of life's history, but the choice between competing phylogenetic methods for the analysis of morphology is unclear. Traditionally, parsimony methods have been favoured but recent studies have shown that these approaches are less accurate than the Bayesian implementation of the Mk model. Here we expand on these findings in several ways: we assess the impact of tree shape and maximum-likelihood estimation using the Mk model, as well as analysing data composed of both binary and multistate characters. We find that all methods struggle to correctly resolve deep clades within asymmetric trees, and when analysing small character matrices. The Bayesian Mk model is the most accurate method for estimating topology, but with lower resolution than other methods. Equal weights parsimony is more accurate than implied weights parsimony, and maximum-likelihood estimation using the Mk model is the least accurate method. We conclude that the Bayesian implementation of the Mk model should be the default method for phylogenetic estimation from phenotype datasets, and we explore the implications of our simulations in reanalysing several empirical morphological character matrices. A consequence of our finding is that high levels of resolution or the ability to classify species or groups with much confidence should not be expected when using small datasets. It is now necessary to depart from the traditional parsimony paradigms of constructing character matrices, towards datasets constructed explicitly for Bayesian methods.

1. Introduction

The fossil record affords the only direct insight into evolutionary history of life on the Earth, but the incomplete preservation and temporal distribution of fossils has long prompted biologists to seek alternative perspectives, such as molecular phylogenies of living species, eschewing palaeontological evidence altogether [1]. However, there is increasing acceptance that analyses of historical diversity cannot be made without phylogenies that incorporate fossil species [2,3] and calibrating molecular phylogenies to time cannot be achieved effectively without recourse to the fossil record [4]. Integrating fossil and living species has become the grand challenge and there has been a modest proliferation of phylogenetic approaches to the analysis of phenotypic data. While conventional parsimony remains the most widely employed method, alternative parsimony [5] and probabilistic [6] models have been developed to better accommodate heterogeneity in

the rate of evolution among characters and across phylogeny. Unfortunately, these competing methods invariably yield disparate phylogenetic hypotheses among which it is difficult to discriminate as the true tree is never known for empirical data.

A number of studies have attempted to establish the efficacy of competing phylogenetic methods using data simulated from known trees [7–9], finding that the probabilistic Mkv model outperforms parsimony methods, among which, conventional equal-weights parsimony (EW-Parsimony) performs best. However, these studies were potentially biased by their experimental design: (i) two of the studies employed a generating tree that was unresolved and, therefore, biased against parsimony methods which recover resolved trees; (ii) these studies did not discriminate between the impact of the probabilistic model and its implementation in a Bayesian framework; (iii) based on single empirical trees, the impact of tree symmetry, which is known to confound phylogeny estimation [10], was not explored; and (iv) only binary characters were considered, whereas empirical datasets are commonly a mixture of binary and multistate characters. Therefore, we compare the performance of EW-Parsimony, implied-weights parsimony (IW-Parsimony), maximum-likelihood and Bayesian implementations of the Mk model, based on datasets with different numbers of characters, comprising binary and multistate characters and simulated on a fully balanced and a maximally imbalanced phylogenetic tree. We find that Bayesian inference outperforms all other methods, while EW-Parsimony performs better than IW-Parsimony, and maximum likelihood performs worst of all. We apply these competing phylogenetic methods to empirical morphological datasets of similar sizes to our simulated datasets and explore the efficacy of the ensuing phylogenetic hypotheses in the light of the conclusions derived from our simulation-based study.

2. Material and methods

(a) Simulation of morphological matrices

We simulated data on two 32-taxon generating trees at the extremes of tree symmetry: one fully asymmetrical and one fully symmetrical (see electronic supplementary material, figure S1). For each tree, we simulated matrices of three sizes: 100, 350 and 1000 characters. We generated matrices using the HKY + Γ Continuous model of molecular substitution, with $\kappa = 2$, the shape (set equal to rate) of the gamma distribution and underlying substitution rate for each replicate sampled from independent and identically distributed exponential distributions with a mean of 1, and character state stationary frequencies fixed as $\pi = [0.2,0.2,0.3,0.3]$. We used a fixed and uneven stationary distribution of nucleotide frequencies to ensure our simulation model did not collapse into the Mk model, as this would bias the analysis in favour of Mk model-based approaches. We simulated 1000 replicate matrices with unique substitution parameters for each tree and each character number, resulting in a total of 6000 matrices. We set two types of character within each matrix, binary and multistate, and we simulated a proportion of 55 binary : 45 multistate characters, based on the mean ratio found in a survey of empirical morphological data matrices [11]. We established binary characters by converting data simulated under the HKY model to R/Y coding (i.e. 0/1): morphological multistate characters were simulated by converting DNA bases to integers.

To ensure that our simulated data are realistic, we generated each set of 1000 unique replicate matrices such that the among-matrix distribution of homoplasy approximated the distribution of empirical homoplasy, characterized by the consistency index (CI), reported by Sanderson & Donoghue [12]. To approximate this distribution of homoplasy, we placed the Sanderson and Donoghue data into quantized bins of CI spanning 0.05, between the empirical bounds of 0.26 and 1.0, and simulated matrices until we matched this expected density per bin (electronic supplementary material, figure S2).

The code used to simulate these data is available in the electronic supplementary material.

(b) Phylogenetic analysis

We analysed the simulated matrices with EW-Parsimony, IW-Parsimony ($k = 2$) and the Mk model [6] under both maximum-likelihood and Bayesian implementations. EW-Parsimony and IW-Parsimony estimation of topology was performed in TNT [13]. We used the Mk + Γ model for maximum-likelihood estimation of topology in RAxML v. 7.2 [14], and Bayesian estimation of topology in MrBayes v. 3.2 [15]. As the approximate likelihood calculation of RAxML may be distant from the true likelihood [16], we conducted a sensitivity test by re-analysing a subset of our data with the likelihood implementation of the Mk model in IQ-tree [17]; both methods gave effectively identical results, indicating results from the likelihood Mkv model are not software specific.

The Mkv model is inappropriate due to the lack of acquisition bias in the simulated data. For maximum-likelihood and Bayesian analyses, we applied the discretized gamma distribution model to account for between-character rate heterogeneity. For Bayesian analyses, the posterior distribution was sampled 1 million times by four chains using the Metropolis-coupled Markov-chain Monte Carlo algorithm with every 100th sample stored, resulting in 10 000 samples; two independent runs were performed for each replicate and the two resulting posterior samples were combined after qualitative assessment of convergence. For parity, we characterized the result of all phylogenetic methods as the majority-rule consensus of resultant tree samples. We did not employ bootstrap methods to measure support for parsimony and likelihood analyses because phenotypic data does not meet the assumption that phylogenetic signal is distributed randomly among characters.

We used the Robinson–Foulds metric [18] to compare the similarity of estimated topologies against their respective generating tree. We also noted the per-node resolution, and the variation of node accuracy across the topology.

(c) Empirical analyses

We analysed four published palaeontological phenotype character matrices that encompass a range of character numbers and a diverse sample of taxa from the Tree of Life [19–22]. We resolved any ambiguities in character coding to their most derived state for each matrix to make analyses compatible across the different phylogenetic methods, facilitating comparison of results. We analysed each matrix by applying the same settings used to analyse our simulated matrices: EW-Parsimony, IW-Parsimony, as well as Bayesian and maximum-likelihood implementations of the Mk model. Empirical morphological matrices are rarely constructed to contain invariant or parsimony uninformative characters. Therefore, the Mkv extension of the Mk model, which uses conditional likelihood to correct for such acquisition biases, is more appropriate than the Mk model for analysis of these empirical data matrices [6].

3. Results

(a) Simulated data

Accuracy is higher for trees inferred from data simulated on a symmetrical topology compared with trees

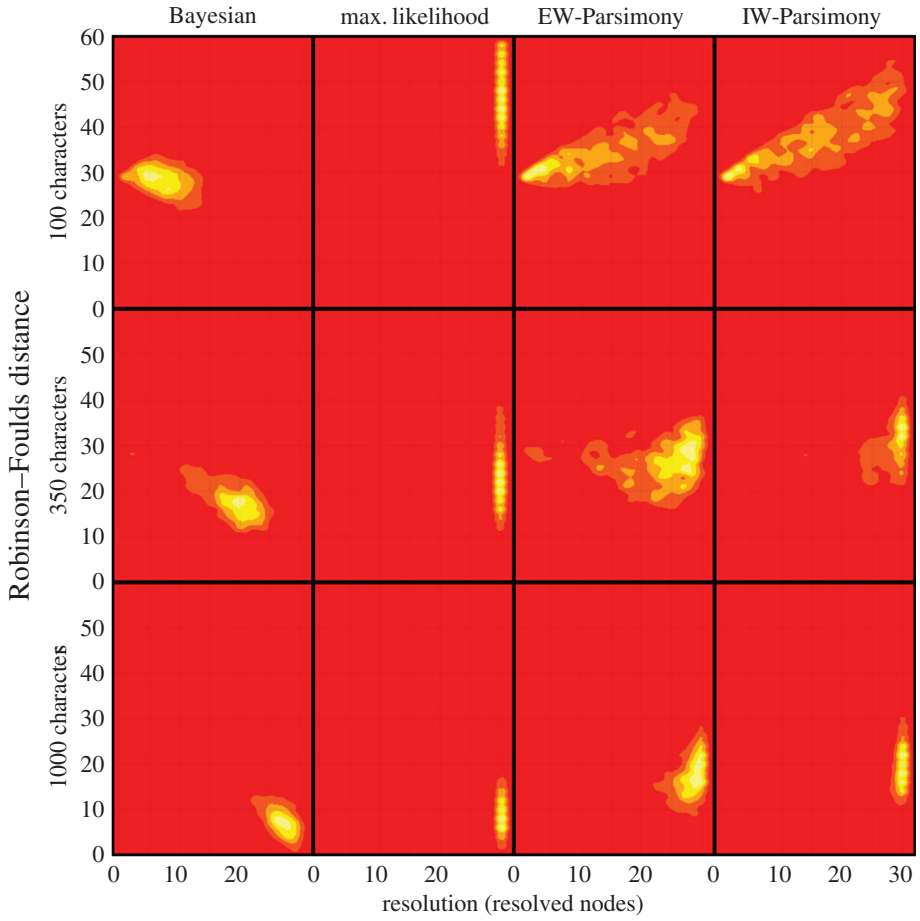


Figure 1. Contour plots of Robinson–Foulds distance against phylogenetic resolution, indicating the higher accuracy of Bayesian implementations against all other methods with data generated on the asymmetrical phylogeny. The spectrum of red to yellow, reflect lower to higher density of trees. As the number of characters increases all methods converge on the correct phylogeny, although Bayesian phylogenies are generally the least resolved. The other methods achieve higher resolution but at a cost of lower accuracy. Data generated on the symmetrical phylogeny shows similar patterns but with much less variance and higher accuracy for all iterations; this lack of variance means point estimates cannot be shown as density estimates. (Online version in colour.)

estimated from data simulated on the asymmetrical topology (cf. figures 2 and 3). Bayesian consensus phylogenies are generally the least well-resolved (figure 1). All methods estimated topologies with greater accuracy as the number of analysed characters increased (figures 2 and 3; electronic supplementary material, table S5–S7). All methods, apart from maximum likelihood, produced phylogenies with greater resolution with higher numbers of characters (figure 1).

For all implementations and dataset sizes, the Bayesian implementation of the Mk model achieves higher accuracy compared with other methods (table 1; figures 1–3). The two parsimony methods achieved the next highest levels of accuracy, EW-Parsimony achieving greater accuracy than IW-Parsimony. Maximum likelihood was the least accurate method for topology reconstruction for both the symmetrical and asymmetrical phylogenies (table 1). The relative accuracy of these phylogenetic methods remains the same across all dataset sizes and the two simulation topologies (table 1; figures 1–3).

Nodes closer to the tips are significantly more accurately reconstructed in the asymmetrical phylogenies across all dataset sizes (table 2 and figure 2; electronic supplementary material, figure S8). In the symmetrical trees, there was no significant correlation between distance from the tips and the accuracy of node reconstruction, except in the maximum-likelihood analysis of 100 characters (figure 2 and table 2).

(b) Empirical phylogenies

Patterns of resolution achieved from the simulated datasets are similar for the empirical datasets. The Bayesian implementation of the Mk model estimates the least resolved phylogenies and maximum likelihood produces fully resolved trees (full trees are shown electronic supplementary material, figure S9–S15).

Kulindroplax, from the Sutton *et al.* [22] dataset, is supported as a crown-mollusc based on maximum likelihood, EW-Parsimony and IW-Parsimony (figure 4*a–d*). The results of the IW-Parsimony analysis are most similar to the original results [22], with *Kulindroplax* resolved as a crown-aplacophoran; maximum-likelihood analysis of the dataset resolved *Kulindroplax* as the stem-aplacophoran. The result of the Bayesian analysis of the dataset is largely unresolved, and *Kulindroplax* is not discriminated as a member of any clade within molluscs or even as a member of total-group Mollusca.

The anthophyte hypothesis (non-monophyletic gymnosperms sister to seed ferns plus angiosperms) recovered by Hilton & Bateman [19] is supported by our EW-Parsimony and maximum-likelihood analyses of their dataset which recovered a paraphyletic seed ferns plus Gnetophyta as sister to angiosperms (figure 4*f,g*); the results of Bayesian and IW-Parsimony analyses of the same dataset contradict the anthophyte hypothesis (figure 4*e,h*). The Bayesian analysis produced a non-monophyletic gymnosperms with the relationships between them and seed ferns unresolved with the exception of

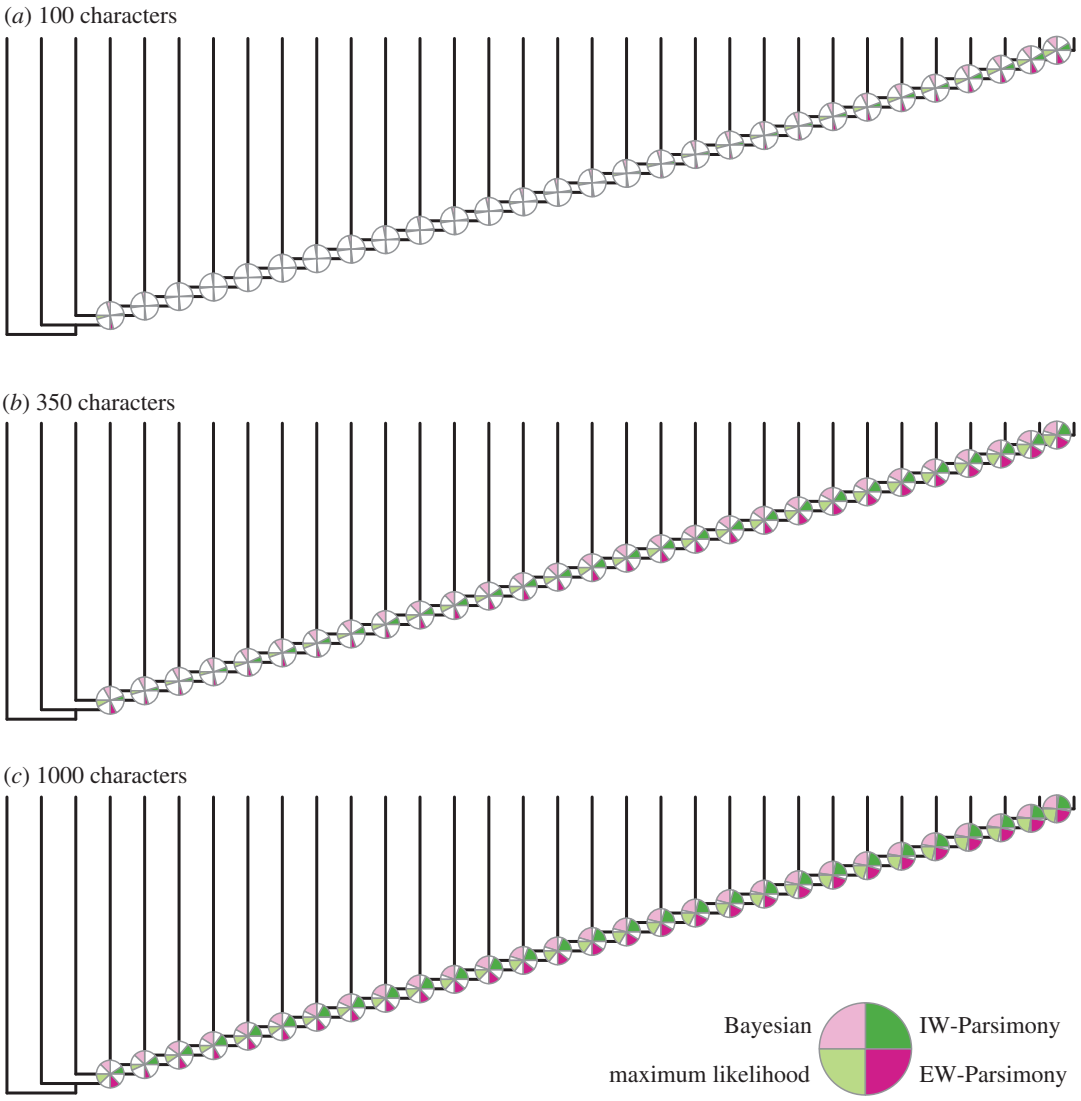


Figure 2. Accuracy of nodes is higher for those closer to the tips in the asymmetrical trees. The percentage of times a node was accurately reconstructed is shown as a proportion of a quarter of a circle in anticlockwise order for Bayesian, maximum likelihood, EW-Parsimony and IW-Parsimony at each node. Accuracy of reconstructions is significantly lower in the 100 character dataset (*a*), and increases in the 350 character (*b*) and 1000 character datasets (*c*). (Online version in colour.)

Bennettitales which resolved as a gnetophyte, and *Caytonia* as sister to the angiosperms.

Analyses of the Luo *et al.* [20] dataset yielded congruent results with the original study, with the placement of *Haramiyavia* outside of crown-Mammalia and multituberculates, although some haramiyids are resolved as crown mammals in the IW-Parsimony analysis (figure 5*a–d*).

Nyasasaurus is recovered as a member of Dinosauria in the maximum likelihood, EW-Parsimony and IW-Parsimony analyses of the dataset from Nesbitt *et al.* [21] (figure 5*e–h*). The Bayesian analysis recovers *Nyasasaurus* in a polytomy with the two major clades of dinosaurs, corroborating the conclusion of Nesbitt *et al.* [21] that, given the data, its precise phylogenetic position is uncertain.

4. Discussion

(a) Simulations indicate that the Bayesian implementation of the Mk model outperforms all other methods and implementations

Previous simulation-based analyses that have attempted to evaluate the performance of likelihood and parsimony-based phylogenetic methods for analysing phenotypic data have found that the probabilistic model performs best [7,8]. However, these studies were biased against parsimony because they employed an unresolved generating tree that is problematic as parsimony methods will attempt to recover a fully resolved tree from the simulated data yielding a non-zero RF distance from the generating tree, even if the two trees are effectively compatible. Furthermore, since previous simulation studies considered the Mk model only within a Bayesian framework, they did not distinguish between the impact of the probabilistic model of character evolution and the statistical framework in which it was implemented.

Our analyses control for these shortcomings of previous simulation studies and show consistently that the Bayesian implementation of the Mk model performs best. In line with previous simulations [8], we found that EW-Parsimony performs better than IW-Parsimony. There is overlap between model performance shown by the distribution of Robinson–Foulds distances (table 1), but there is reason to have different degrees of confidence in the models; only the Bayesian implementation produces a relatively small distribution of tree performance compared with the large tails signifying worse performance in the two parsimony methods (table 1). We also found that the Bayesian implementation of the Mk model outperforms the

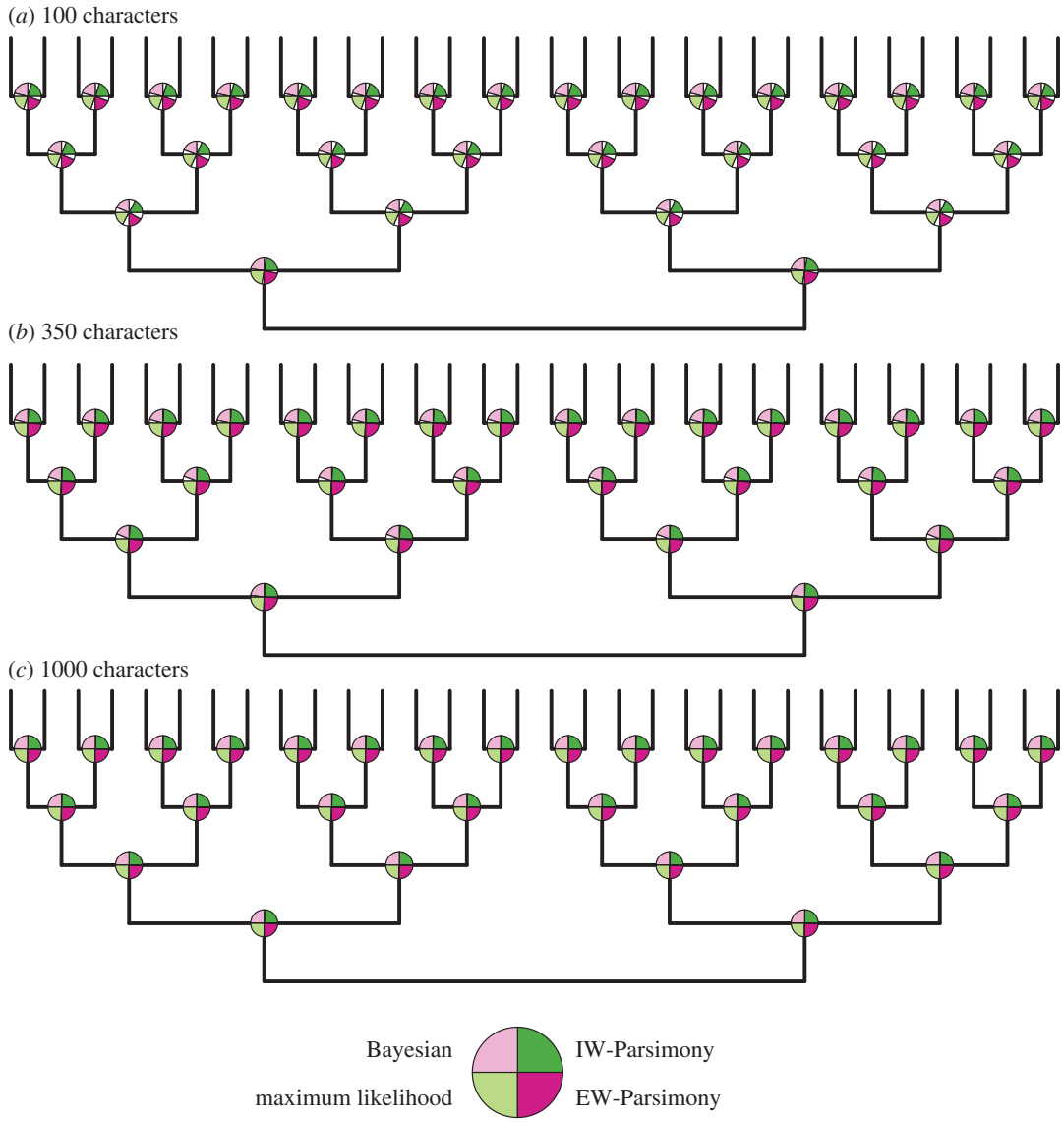


Figure 3. Accuracy of nodes is high for all nodes in the symmetrical phylogeny. The percentage of times a node was accurately reconstructed is shown as a proportion of a quarter of a circle in anticlockwise order for Bayesian, maximum likelihood, EW-Parsimony and IW-Parsimony at each node. Accuracy of reconstructions is high in each dataset size, but there is a non-significant increase in accuracy as dataset size increases (*a*–*c*). (Online version in colour.)

Table 1. Bayesian approaches produce the most accurate trees for all character sets. Mean and range (in brackets) of Robinson–Foulds distances are lower for topologies estimated using Bayesian methods for both the symmetrical and asymmetrical generating tree. Maximum likelihood is the generally the most inaccurate method for the symmetrical generating tree, and implied weights parsimony performs worst for the asymmetrical generating tree.

	equal weights parsimony	implied weights parsimony	maximum likelihood	Bayesian
asymmetrical generating phylogeny				
100	34.89 (22–56)	37.85 (22–56)	45.84 (20–58)	28.1 (18–39)
350	26.57 (11–51)	29.2 (12–51)	26.49 (6–58)	19.21 (7–35)
1000	17.82 (3–40)	19.16 (2–33)	11.94 (0–58)	9.34 (0–31)
symmetrical generating phylogeny				
100	8.08 (0–33)	9.29 (0–29)	10.1 (0–58)	7.51 (0–29)
350	1.33 (0–28)	1.43 (0–28)	1.8 (0–52)	1.2 (0–28)
1000	0.32 (0–26)	0.31 (0–26)	0.51 (0–52)	0.31 (0–26)

maximum-likelihood implementation, indicating that it is not merely the probabilistic transition model that outperforms parsimony methods, but the implementation of the Mk model within a Bayesian statistical framework. Indeed, the maximum-likelihood implementation of the Mk model was the worst-performing method, worse even than IW-Parsimony. In part, the poor performance of the maximum-likelihood-Mk method is because we did not capture phylogenetic uncertainty

Table 2. *p*-Values from Spearman's rank correlation between the percentage of nodes being accurately reconstructed and their distance from the root. Nodes closer to the tips are significantly more likely to be accurately reconstructed in asymmetrical trees but this is not generally true for symmetrical phylogenies.

	asymmetrical tree	symmetrical tree
MB 100	<0.001	0.09919
maximum likelihood 100	<0.001	0.027295
EW 100	<0.001	0.106712
IW 100	<0.001	0.092736
MB 350	<0.001	0.638242
maximum likelihood 350	<0.001	0.057809
EW 350	<0.001	0.19683
IW 350	<0.001	0.148108
MB 1000	<0.001	0.256976
maximum likelihood 1000	<0.001	0.085987
EW 1000	<0.001	0.179186
IW 1000	<0.001	0.287058

associated with this phylogenetic method. This is normally achieved in analyses of molecular datasets through bootstrapping methods, but these are inappropriate for the analysis of phenotypic data as the basic methodological assumption, that the phylogenetic signal is randomly distributed across sites (characters), is not true for morphological data.

However, irrespective of the phylogenetic method used, dataset size correlated positively with both phylogenetic accuracy and resolution, diminishing differences in the relative performance of the competing phylogenetic methods. All phylogenetic methods also performed best when attempting to recover a symmetrical target tree; all methods found recovery of asymmetrical trees challenging and phylogenetic accuracy diminished from tip to root. The impact of tree topology is of particular concern since empirical phylogenetic trees are invariably asymmetric [23], and trees of fossil species are infamous for their asymmetry [24,25]. However, there is a broad spectrum of tree symmetry, with fully symmetric and fully asymmetric trees representing end-members. Palaeontological trees with the dimensions used in our simulations are typically far from the fully asymmetric pectinate-generating tree we employed (Ic = ~0.4 for 32 species) [25]. Furthermore, the asymmetry of many palaeontological trees is often a representational artefact of attempting to summarize character evolution, or an analytic artefact of analysing the relationships among diverse clades based on representative species or higher taxa [26]. Thus, the challenge of recovering trees of extinct taxa may not be as great as a simplistic interpretation of our results might suggest.

(b) Analyses of empirical data bear out conclusions based on simulations

Maximum-likelihood, IW-Parsimony and EW-Parsimony methods of the simulated datasets commonly identify a single optimal tree, but the differences between the optimal trees derived from these methods provides no confidence that any one of the inferred topologies is accurate with reference to the placement of a taxon of interest. This view is corroborated by our reanalysis of empirical datasets which recovered poorly resolved trees using the Bayesian implementation of the Mk model, and in a number of instances, indicate that the conclusions drawn in the corresponding original studies are not supported by the data.

In an extreme example, our re-analyses of the dataset published by Sutton *et al.* [22], which attempted to demonstrate a crown-aplacophoran mollusc affinity for *Kulindroplax*, yielded disparate hypotheses of affinity. EW-Parsimony and IW-Parsimony recovered the published result, while maximum likelihood recovered *Kulindroplax* as a stem-aplacophoran, and Bayesian could not discriminate *Kulindroplax* as a total-group mollusc (figure 4*a*). This poor resolution is unlikely to be a result of poor fossil evidence but, rather, the lack of discriminatory power in the small character matrix. Among the analyses of the dataset from Hilton & Bateman [19], we recovered some of the principal competing topologies that have featured in debate over the affinity of seed plants in past decades. However, the Bayesian analysis of the dataset recovered a topology that is largely unresolved in terms of the relationships among key clades. This suggests that the available data are insufficient to discriminate among the competing hypotheses, and this long-standing debate is largely an artefact of the false resolution of parsimony methods.

Bayesian analyses need not overturn the results from previous analyses based on deterministic phylogenetic methods like EW-Parsimony, IW-Parsimony and maximum likelihood. A phylogenetic position for haramiyids, outside crown-Mammalia, is corroborated by our Bayesian analysis of the dataset from Luo *et al.* [20]—in contrast with the crown-Mammalia affinity recovered for some haramiyids through IW-Parsimony analysis of the same data (figure 5*d*). Similarly, *Nyasasaurus* was posited as the earliest dinosaur, and this conclusion is supported by the Bayesian analyses (figure 5*e*) although this is not supported by EW-Parsimony, IW-Parsimony and maximum-likelihood analyses (figure 5*f–h*). However, the Bayesian analysis is more robust in expressing the phylogenetic ambiguity identified by the original authors [19], as *Nyasasaurus* falls in a polytomy alongside the two major clades of dinosaurs.

Some of the differences between methods may simply reflect the dimensions of the dataset. The two datasets that cannot resolve relationships under Bayesian inference and exhibit significant topological discordance among phylogenetic methods [19,22] are both comparatively small (34 taxa, 48 characters and 48 taxa, 82 characters). These both fall within the scope of simulated datasets that yield low resolution from the Bayesian method and, from other phylogenetic methods, high resolution but low accuracy (figure 1). The two empirical datasets that yield trees with greater congruence from the different phylogenetic methods, are both larger: Luo (114 taxa, 497 characters) and Nesbitt (82 taxa, 413 characters). The size of these matrices is comparable with our simulation results in which we see marked increases in topological accuracy and agreement between methods (figure 1, between 350 and 1000 characters).

(c) Implications for phylogenetic analysis of phenotypic data

The results of our simulation studies indicate that the cadre of phylogenetic hypotheses generated from phenotypic data

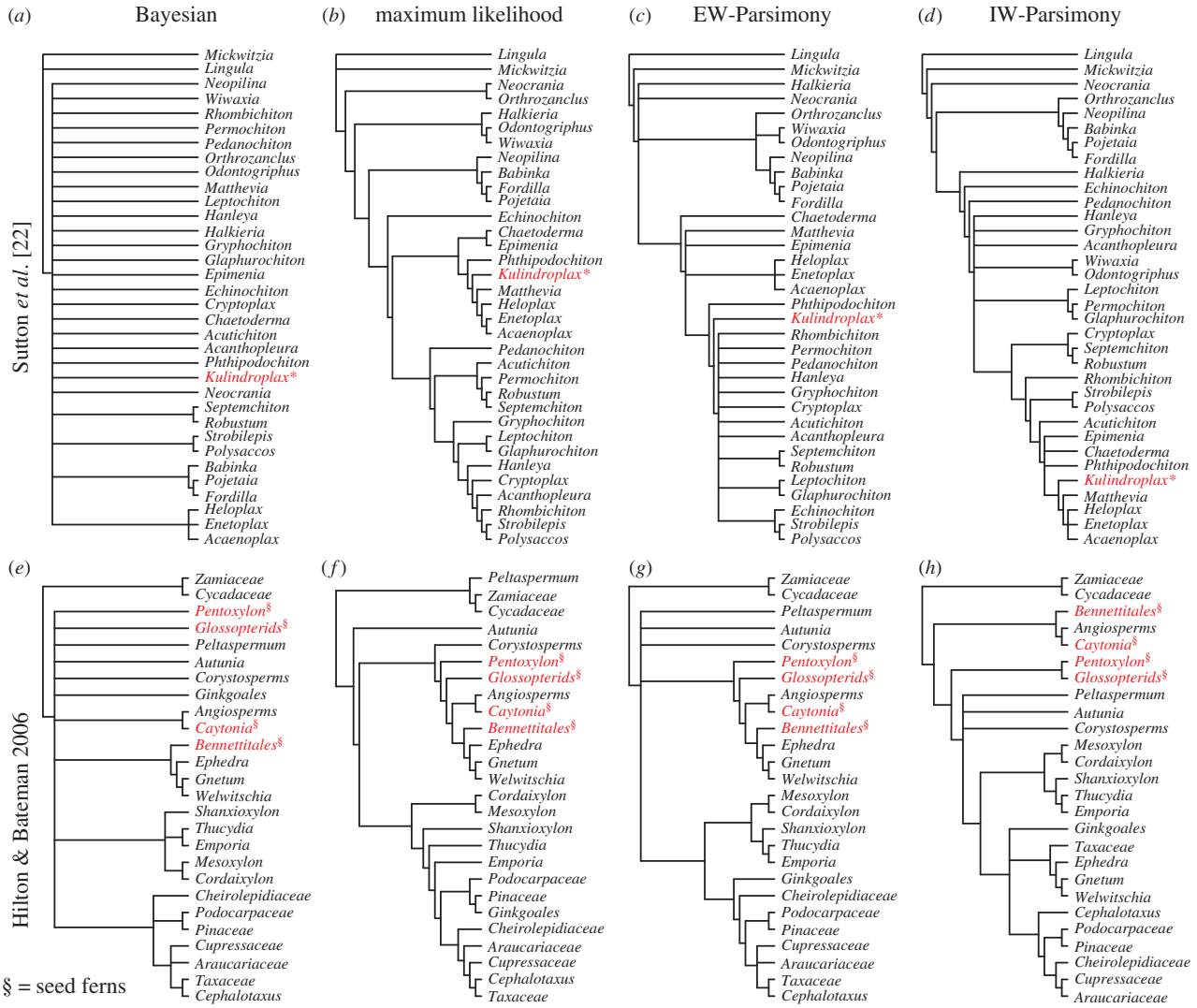


Figure 4. Alternative phylogenetic reconstruction methods alter our understanding of evolution with empirical matrices. However, the relationship of fossil seed ferns from Hilton & Bateman [19] is changed according to implementation (*a–d*), although *Caytonia* remains as sister to angiosperms in all analyses. Alternative analyses change the taxonomic affinity of *Kulindroplax* from Sutton *et al.* [22] (*e–h*). (Online version in colour.)

using parsimony methods require reassessment using the Bayesian implementation of the Mk model. It is likely that many evolutionary interpretations are contingent on precise but inaccurate phylogenetic hypotheses. In this undertaking, it is important that the implications of our simulation studies are considered in the design of phylogenetic studies.

Firstly, phylogenies of fossils tend towards strong asymmetries [25] and, like all phylogenetic methods, Bayesian inference struggles with the recovery of deep nodes within asymmetric trees. Therefore, it is important that outgroups are sampled extensively, ensuring that contentious in-group relationships are closer to the tips, where topological accuracy is highest. Further, in-group lineages should be sampled in a manner that does not accentuate tree asymmetry.

Secondly, phylogenetic accuracy and resolution correlates positively with the relative dimensions of the dataset. Accordingly, phylogenetic resolution or certainty should not be expected from cladistic analyses of small morphological datasets (i.e. those around 100 characters or fewer), particularly if they include fossils. There are finite limits to the number of available phylogenetically informative characters [27] and, for well-studied clades, it may be perceived that these phylogenetically informative characters have already been found. However, it is important to note that the concept of phylogenetic informativeness is different within a likelihood versus a parsimony framework: in parsimony characters that undergo few changes are prized in favour of homoplastic characters. Under the likelihood model, branch length, informed by the number of character changes, contributes to topology estimation. Thus, traditionally 'bad' phylogenetic characters (those exhibiting homoplasy) may find utility in expanding the dimensions of phenotypic character matrices as long as homoplasy falls within the limits that the model can accommodate. In a Bayesian framework, this can be tested using posterior predictive tests of model adequacy (e.g. [28]).

Finally, we may need to alter our expectations to anticipate less well-resolved but more accurate phylogenetic hypotheses, which will both constrain and guide research. Greater resolution may be found by generating matrices suited to likelihood- rather than parsimony-based phylogenetic methods. However, we must also come to terms with the prospect that for some groups of organisms, or their fossil remains, there may be insufficient data. As such, their evolutionary relationships might not therefore be resolvable using morphological data alone and, if they are fossils, their evolutionary significance may never be realized. Nevertheless, resolving phylogenies is not the end game for evolutionary biology.

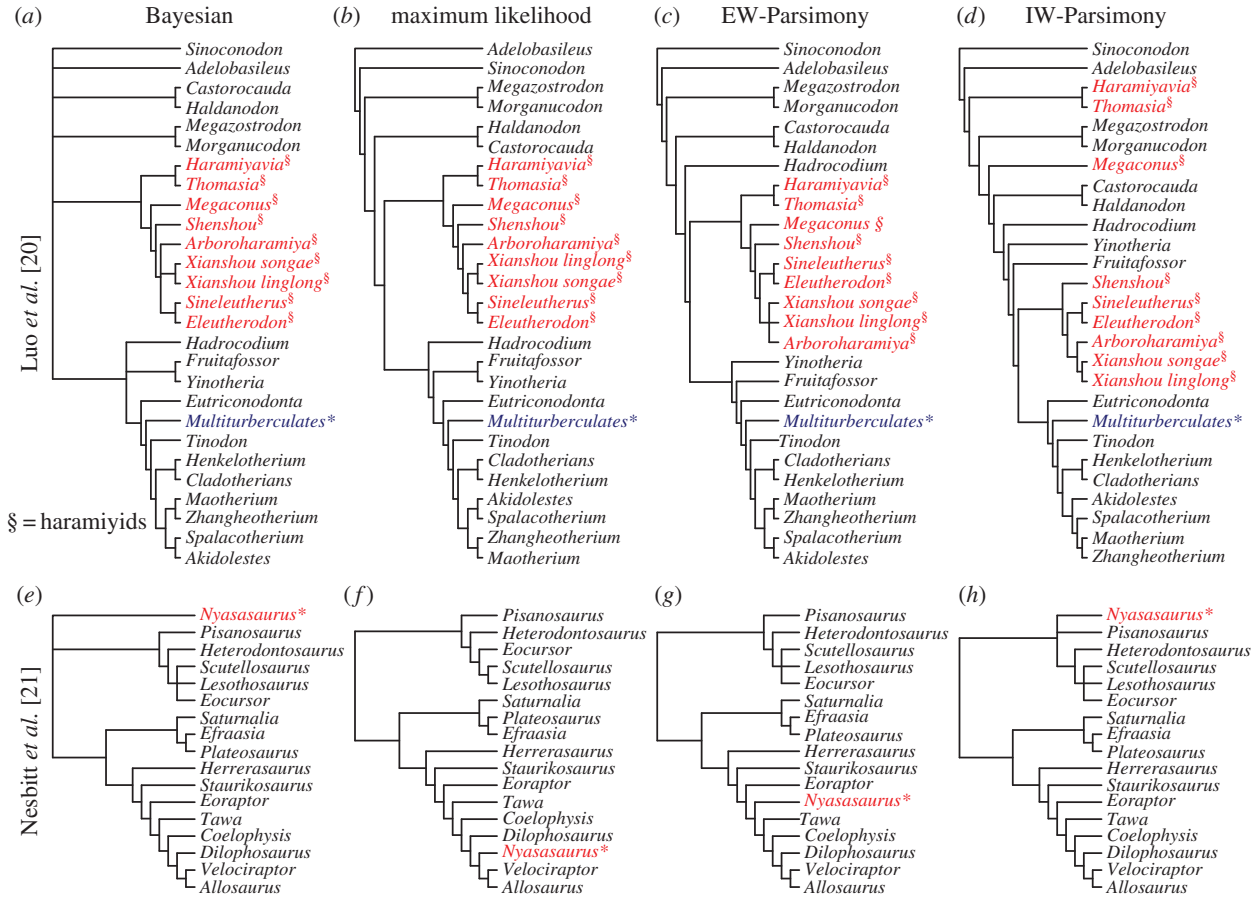


Figure 5. Alternative phylogenetic reconstruction methods produce generally congruent reconstructions of evolution with empirical matrices. For Luo *et al*. [20], the relationship between the haramiyids and multituberculates is largely unchanged across analyses (*a–d*). IW-Parsimony (*g*) and Bayesian analyses place *Nyasasaurus* as close to the earliest dinosaur (*e*) and IW-Parsimony places it close to the earliest diverging taxa (*g*), but EW-Parsimony and maximum likelihood place the taxa as a derived member of Dinosauria (*f*,*h*). (Online version in colour.)

Incompletely resolved trees can still be used as a basis for investigating interesting macroevolutionary questions, and methods exist for incorporating tree uncertainty in phylogenetic comparative methods (e.g. [29]).

5. Conclusion

A growing consensus shows that the Bayesian Mk model is the most accurate method of phylogenetic reconstruction, and here we show that this remains true across dramatically different tree shapes, when analysing datasets composed of both multistate and binary characters, and when compared with maximum-likelihood estimation using the Mk model. We recommend that Bayesian implementations of the Mk model should become the default method for phylogenetic analyses of cladistic morphological datasets, and we should expect low levels of resolution with small datasets. As parsimony methods appear to be less effective than probabilistic approaches, it may be necessary to alter data collection practices by moving away from choosing a selection of characters that undergo few changes, and moving towards scoring all possible characters from the available taxa irrespective of their expected homoplasy.

Data accessibility. Supplementary figures and the code used to simulate the data used in this publication can be accessed in the electronic supplementary material.

Authors' contributions. All authors contributed to the design of the study; M.N.P. and J.E.O.R. led the analyses; interpretation of results and writing was led by M.N.P., J.E.O.R., P.C.J.D. and D.P., though all authors contributed to the interpretation of results and the writing of the manuscript.

Competing interests. We have no competing interests.

Funding. This research was funded by NERC (NE/L501554/1 to J.E.O.R. and L.A.P.; NE/K500823/1 to M.N.P.; NE/L002434/1 to J.F.; NE/N003438/1 to P.C.J.D.), BBSRC (BB/N000919/1 to P.C.J.D.), the University of Bristol (STaR scholarship to A.R.T.), Royal Society Wolfson Research Merit Award (P.C.J.D.) and the John Templeton Foundation (43915 to D.P. and L.H.).

Acknowledgements. We thank the other members of the Bristol Palaeobiology research group for discussion; Rob Asher (Cambridge) and Thomas Guillerme for comments on the draft manuscript. We also thank April Wright and an anonymous reviewer for their help in improving the manuscript.

References

1. Harvey P, May R, Nee S. 1994 Phylogenies without fossils. *Evolution*. **48**, 523–529. (doi:10.2307/2410466)

2. Rabosky DL. 2010 Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824. (doi:10.1111/j.1558-5646.2009.00926.x)

3. Losos JB *et al.* 2013 Evolutionary biology for the 21st century. *PLoS Biol*. **11**, e1001466. (doi:10.1371/journal.pbio.1001466)
4. dos Reis M, Donoghue PCJ, Yang Z. 2016 Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet*. **17**, 1–10. (doi:10.1038/nrg.2015.8)
5. Goloboff PA, Carpenter JM, Arias JS, Miranda-Esquivel DR. 2008 Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* **24**, 758–773. (doi:10.1111/j.1096-0031.2008.00209.x)
6. Lewis PO. 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol*. **50**, 913–925. (doi:10.1080/106351501753462876)
7. Wright AM, Hillis DM. 2014 Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* **9**, e109210. (doi:10.1371/journal.pone.0109210)
8. O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PCJ. 2016 Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biol. Lett*. **12**, 20160081. (doi:10.1098/rsbl.2016.0081)
9. Congreve CR, Lamsdell JC. 2016 Implied weighting and its utility in palaeontological datasets: a study using modelled phylogenetic matrices. *Palaeontology* **59**, 447–462. (doi:10.1111/pala.12236)
10. Holton TA, Wilkinson M, Pisani D. 2014 The shape of modern tree reconstruction methods. *Syst. Biol*. **63**, 436–441. (doi:10.1093/sysbio/syt103)
11. Guillerme T, Cooper N. 2016 Effects of missing data on topological inference using a total evidence approach. *Mol. Phylogenet. Evol*. **94**, 146–158. (doi:10.1016/j.ympev.2015.08.023)
12. Sanderson MJ, Donoghue M. 1996 *The relationship between homoplasy and the confidence in a phylogenetic tree*. San Diego, CA: Academic Press.
13. Goloboff PA, Farris S, Nixon K. 2008 TNT, a free programm for phylogenetic analysis. *Cladistics* **24**, 774–786. (doi:10.1111/j.1096-0031.2008.00217.x)
14. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
15. Ronquist F *et al.* 2012 Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol*. **61**, 539–542. (doi:10.1093/sysbio/sys029)
16. Wright AM, Lyons KM, Brandley MC, Hillis DM. 2015 Which came first: the lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. *J. Exp. Zool. B. Mol. Dev. Evol*. **324**, 504–516. (doi:10.1002/jez.b.22642)
17. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol*. **32**, 268–274. (doi:10.1093/molbev/msu300)
18. Robinson DF, Foulds LR. 1981 Comparison of phylogenetic trees. *Math. Biosci*. **53**, 131–147. (doi:10.1016/0025-5564(81)90043-2)
19. Hilton J, Bateman RM. 2006 Pteridosperms are the backbone of seed-plant phylogeny. *J. Torrey Bot. Soc*. **133**, 119–168. (doi:10.3159/1095-5674)
20. Luo ZX, Gatesy SM, Jenkins FA, Amaral WW, Shubin NH. 2015 Mandibular and dental characteristics of Late Triassic mammaliaform *Haramiyavia* and their ramifications for basal mammal evolution. *Proc. Natl Acad. Sci. USA* **112**, E7101–E7109. (doi:10.1073/pnas.1519387112)
21. Nesbitt SJ, Barrett PM, Werning S, Sidor CA, Charig AJ. 2013 The oldest dinosaur? A Middle Triassic dinosauriform from Tanzania. *Biol. Lett*. **9**, 20120949. (doi:10.1098/rsbl.2012.0949)
22. Sutton MD, Briggs DEG, Siveter DJ, Siveter DJ, Sigwart JD. 2012 A Silurian armoured aplacophoran and implications for molluscan phylogeny. *Nature* **490**, 94–97. (doi:10.1038/nature11328)
23. Mooers AO, Heard SB. 1997 Inferring evolutionary process from phylogenetics tree shape. *Q. Rev. Biol*. **72**, 31–54. (doi:10.1086/419657)
24. Shao KT, Sokal RR. 1990 Tree balance. *Syst. Zool*. **39**, 266–276. (doi:10.1007/s13398-014-0173-7.2)
25. Harcourt-Brown K, Pearson P, Wilkinson M. 2001 The imbalance of paleontological trees. *Paleobiology* **27**, 188–204. (doi:10.1666/0094-8373(2001)027<0188:TIOPT>2.0.CO;2)
26. Panchen A. 1982 The use of parsimony in testing phylogenetic hypotheses. *Zool. J. Linn. Soc*. **74**, 305–328. (doi:10.1111/j.1096-3642.1982.tb01154.x)
27. Scotland RW, Olmstead RG, Bennett JR. 2003 Phylogeny reconstruction: the role of morphology. *Syst. Biol*. **52**, 539–548. (doi:10.1080/10635150390223613)
28. Tarver JE *et al.* 2016 The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol*. **8**, 330–334. (doi:10.1093/gbe/evv261)
29. Healy K *et al.* 2014 Ecology and mode-of-life explain lifespan variation in birds and mammals. *Proc R. Soc. B* **281**, 20140298. (doi:10.1098/rspb.2014.0298)

The timescale of early land plant evolution

Jennifer L. Morris[a,1], Mark N. Puttick[a,b,1], James W. Clark[a], Dianne Edwards[c], Paul Kenrick[b], Silvia Pressel[d], Charles H. Wellman[e], Ziheng Yang[f,g], Harald Schneider[a,d,h,2], and Philip C. J. Donoghue[a,2]

[a]School of Earth Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom; [b]Department of Earth Sciences, Natural History Museum, London SW7 5BD, United Kingdom; [c]School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10, United Kingdom; [d]Department of Life Sciences, Natural History Museum, London SW7 5BD, United Kingdom; [e]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom; [f]Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom; [g]Radcliffe Institute for Advanced Studies, Harvard University, Cambridge, MA 02138; and [h]Center of Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan 666303, China

Edited by Peter R. Crane, Oak Spring Garden Foundation, Upperville, VA, and approved January 17, 2018 (received for review November 10, 2017)

Establishing the timescale of early land plant evolution is essential for testing hypotheses on the coevolution of land plants and Earth's System. The sparseness of early land plant megafossils and stratigraphic controls on their distribution make the fossil record an unreliable guide, leaving only the molecular clock. However, the application of molecular clock methodology is challenged by the current impasse in attempts to resolve the evolutionary relationships among the living bryophytes and tracheophytes. Here, we establish a timescale for early land plant evolution that integrates over topological uncertainty by exploring the impact of competing hypotheses on bryophyte–tracheophyte relationships, among other variables, on divergence time estimation. We codify 37 fossil calibrations for Viridiplantae following best practice. We apply these calibrations in a Bayesian relaxed molecular clock analysis of a phylogenomic dataset encompassing the diversity of Embryophyta and their relatives within Viridiplantae. Topology and dataset sizes have little impact on age estimates, with greater differences among alternative clock models and calibration strategies. For all analyses, a Cambrian origin of Embryophyta is recovered with highest probability. The estimated ages for crown tracheophytes range from Late Ordovician to late Silurian. This timescale implies an early establishment of terrestrial ecosystems by land plants that is in close accord with recent estimates for the origin of terrestrial animal lineages. Biogeochemical models that are constrained by the fossil record of early land plants, or attempt to explain their impact, must consider the implications of a much earlier, middle Cambrian–Early Ordovician, origin.

plant | evolution | timescale | phylogeny | Embryophyta

The establishment of plant life on land is one of the most significant evolutionary episodes in Earth history. Terrestrial colonization has been attributed to a series of major innovations in plant body plans, anatomy, and biochemistry that impacted increasingly upon global biogeochemical cycles through the Paleozoic. In some models, an increase in biomass over the continents, firstly by cryptogamic ground covers followed by larger vascular plants, enhanced rates of silicate weathering and carbon burial that drove major perturbations in the long-term carbon cycle (1, 2), resulting in substantial drops in atmospheric CO_2 levels (3–6) (but see ref. 7) and increased oxygenesis (8). It also led to new habitats for animals (9) and fungi (10), major changes to soil types (11), and sediment stability that influenced river systems and landscapes (12). Attempts at testing these hypotheses on the coevolution of land plants (embryophytes) and the Earth System have been curtailed by a lack of consensus on the relationships among living plants, the timescale of their evolution, and the timing of origin of key body plan innovations (13). Although the megafossil record provides unequivocal evidence of plant life on land, the early fossil record is too sparse and biased by the nonuniformity of the rock record (13) to directly inform the timing and sequence of character acquisition in the assembly of plant body plans. Therefore, in attempting to derive a timescale for phytoterrestrialization of the planet, we have no recourse but to molecular clock methodology, employing the known fossil record to calibrate and constrain molecular evolution to time. Unfortunately, the relationships among the four principal lineages of land plants, namely, hornworts, liverworts, mosses, and tracheophytes, are unresolved, with almost every possible solution currently considered viable (14). In attempting to establish a robust timeline of land plant evolution, here we explore the impact of these conflicting phylogenetic hypotheses on divergence time estimates of key embryophyte clades.

Early morphology-based cladistic analyses of extant land plants suggested that the bryophytes are paraphyletic, but yielded conflicting topologies (15–17). Molecular phylogenies have been no more certain, with some analyses supporting liverworts as the sister to all other land plants (18), with either mosses (19–21) (Fig. 1*F*), hornworts (22–27) (Fig. 1*E*), or a moss–hornwort clade (28) (Fig. 1*G*) as the sister group to the vascular plants. Variants on these topologies have been suggested, such as a liverwort–moss clade as the sister group to the remaining land plants (29) (Fig. 1*D*). More recently, the debate has concentrated upon two hypotheses: hornworts as the sister to all other land plants (14, 30–34) (Fig. 1*B*) or monophyletic bryophytes sister to the tracheophytes (14, 35, 36) (Fig. 1*A*). Transcriptome-level datasets support both

Significance

Establishing the timescale of early land plant evolution is essential to testing hypotheses on the coevolution of land plants and Earth's System. Here, we establish a timescale for early land plant evolution that integrates over competing hypotheses on bryophyte–tracheophyte relationships. We estimate land plants to have emerged in a middle Cambrian–Early Ordovocian interval, and vascular plants to have emerged in the Late Ordovician–Silurian. This timescale implies an early establishment of terrestrial ecosystems by land plants that is in close accord with recent estimates for the origin of terrestrial animal lineages. Biogeochemical models that are constrained by the fossil record of early land plants, or attempt to explain their impact, must consider a much earlier, middle Cambrian–Early Ordovician, origin.

Author contributions: D.E., P.K., S.P., C.H.W., Z.Y., H.S., and P.C.J.D. designed research; J.L.M., M.N.P., J.C., H.S., and P.C.J.D. performed research; J.L.M., M.N.P., J.C., D.E., P.K., S.P., C.H.W., Z.Y., H.S., and P.C.J.D. analyzed data; and J.L.M., M.N.P., D.E., P.K., S.P., C.H.W., Z.Y., H.S., and P.C.J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All input trees and alignments are available on Figshare (https://dx.doi.org/10.6084/m9.figshare.5573032).

[1]J.L.M. and M.N.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: Phil.Donoghue@bristol.ac.uk or harald@xtbg.ac.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719588115/-/DCSupplemental.

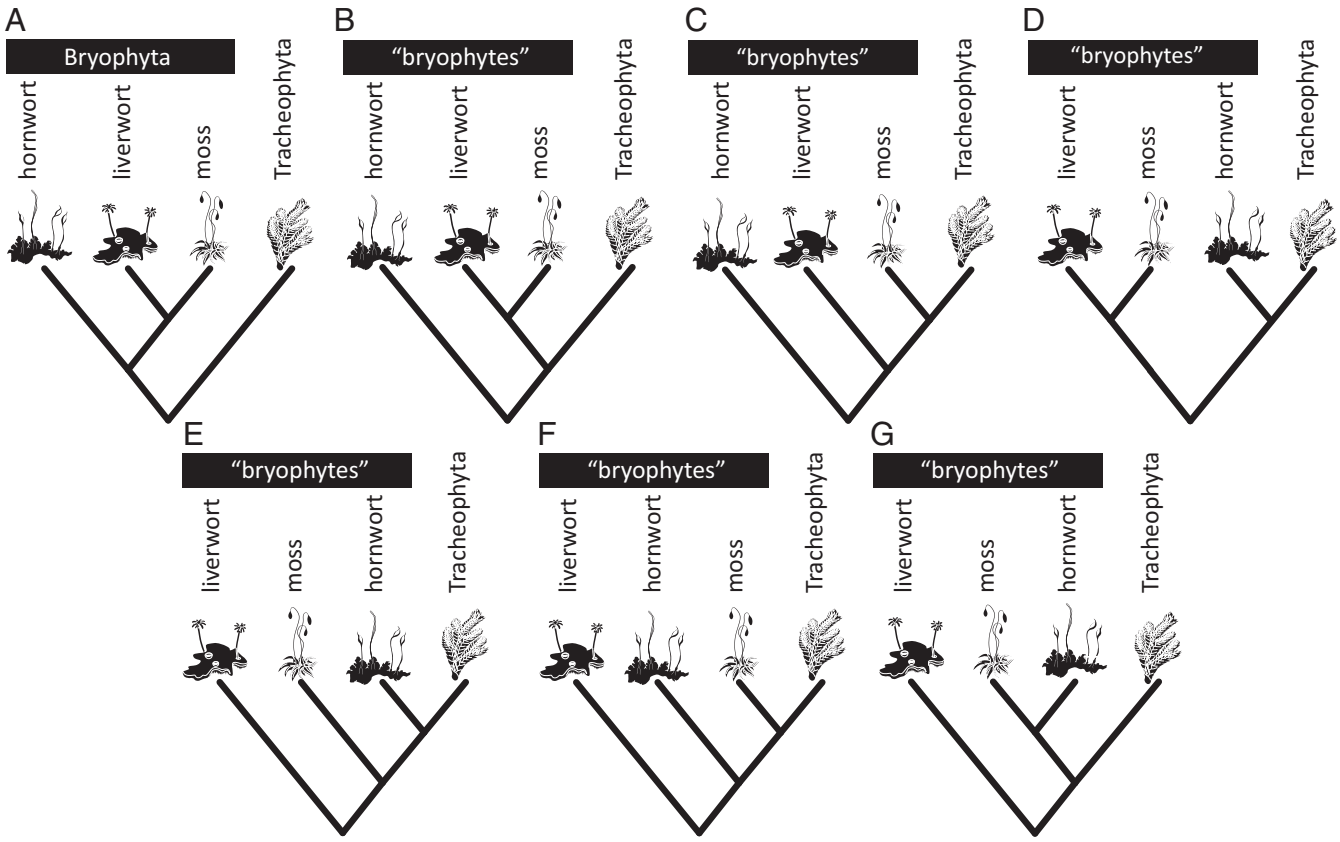


Fig. 1. The seven alternative hypotheses considered in the dating analyses. (*A*) Monophyletic bryophytes; (*B*) liverwort–moss sister clade to tracheophytes; (*C*) mosses, liverworts, and hornworts as successive sister lineages to tracheophytes; (*D*) a moss–liverwort sister clade to other embryophytes; (*E*) hornworts, mosses, and liverworts as successive sister lineages to tracheophytes; (*F*) mosses, hornworts, and liverworts as successive sister lineages to tracheophytes; and (*G*) a moss–hornwort sister clade to tracheophytes.

topologies (14), but sequence heterogeneity makes inferring relationships among these early land plants difficult (36).

Here we attempt to establish a timescale of early land plant evolution that integrates over the contested topological relationships among bryophytes and tracheophytes. To achieve this, we constructed 37 fossil calibrations with minimum and soft maximum constraints, following best practice (37). This requires that calibrations are established on the basis of (*i*) a specific fossil specimen reposited in a publicly accessible collection, (*ii*) an apomorphy-based justification of clade assignment, (*iii*) reconciliation of morphological and molecular phylogenetic context of clade assignment, (*iv*) geographic and

Table 1. Summary of the analyses performed employing the seven alternative hypotheses, removal of the embryophyte constraints, and trimming dataset size

	Node distribution									
	Uniform									
Dataset	850,000	1.7 million	435,000	290,000	19,000	2,000	850,000	1.7 million[†]	Skew-t 850,000	Cauchy 850,000
Dataset no.										
A Monophyletic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
B Hornworts–sister	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
C Hornworts–liverworts–mosses	✓	X	X	X	X	X	X	X	X	X
D Liverworts–mosses–sister	✓	X	X	X	X	X	X	X	X	X
E Liverworts–mosses–hornworts	✓	X	X	X	X	X	X	X	X	X
F Liverworts–hornworts–mosses	✓	X	X	X	X	X	X	X	X	X
G Liverworts–sister	✓	X	X	X	X	X	X	X	X	X
Monophyletic (embryophytes only)	X	✓	X	X	X	X	X	X	X	X
Hornworts–sister (embryophytes only)	X	✓	X	X	X	X	X	X	X	X
Monophyletic (*Chara*–embryophytes)	X	✓	X	X	X	X	X	X	X	X
Hornworts–sister (*Chara*–embryophytes)	X	✓	X	X	X	X	X	X	X	X

All input topologies are based on the 290,718-nucleotide dataset, except for the *Chara*-embryophytes topology, which is based on the likelihood phylogeny of 1.7 million nucleotides.

[†]A correlated model was used to estimate substitution rates on branches rather than the uncorrelated model used in all other analyses.

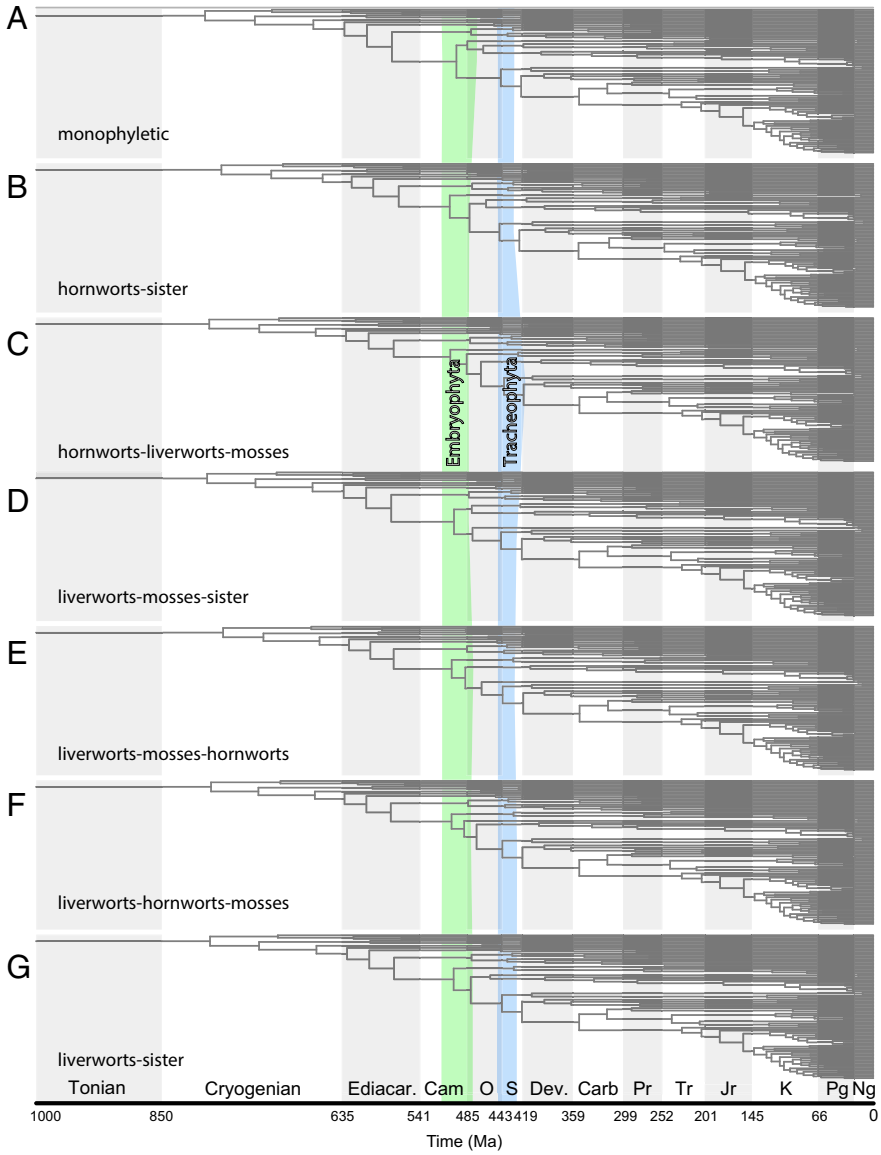


Fig. 2. Age estimates for the seven topologies used in analyses, highlighting the 95% HPD age uncertainty for embryophytes and tracheophytes. Age estimates are shown for (*A*) monophyletic bryophytes, (*B*) hornworts–sister, (*C*) hornworts–liverworts–mosses, (*D*) liverworts–mosses–sister, (*E*) liverworts–mosses–hornworts, (*F*) liverworts–hornworts–mosses, and (*G*) liverworts–sister.

stratigraphic provenance, and (*v*) justification of geochronological age interpretation. Thus defined, these calibrations were combined with existing genetic data (14) in a Bayesian relaxed molecular clock analysis in which we also explored the impact of genetic dataset size and competing calibration strategies, as well as alternative substitution models, on divergence time estimates (Table 1). We find that topology and dataset size have minimal impact on age estimates, but slightly more variance in clade age estimates occurred when using alternative calibration strategies. We conclude that embryophytes emerged within a middle

Table 2. The 95% HPD age estimates for of embryophytes and tracheophytes from divergence time analyses using the seven alternative topologies

Topology	Embryophytes, Ma	Tracheophytes, Ma
Dataset no.		
A Monophyletic	514.8–473.5	450.8–431.2
B Hornworts–sister	515.2–482.1	450.8–430.4
C Hornworts–liverworts–mosses	515.2–483.3	450.7–419.3
D Liverworts–mosses–sister	514.9–477.7	450.8–431.1
E Liverworts–mosses–hornworts	515.1–480.8	450.7–427.9
F Liverworts–hornworts–mosses	515.1–483.2	450.7–428.5
G Liverworts–sister	514.9–478.4	450.8–428.2

Table 3. The 95% HPD age estimates for named nodes in the analyses using the two main topologies of early land plants (monophyletic, hornworts–sister)

Clade	Monophyletic, Ma	Hornworts–sister, Ma
Viridiplantae	972.4–669.9	968.0–676.7
Streptophyta	890.9–629.1	875.4–637.4
Embryophyta	514.8–473.5	515.2–482.1
Bryophytes	506.4–460.3	N/A
Marchantiophyta	443.6–405.3	442.0–405.3
Marchantiopsida	354.9–228.0	357.9–228.0
Bryophyta	448.6–344.3	443.0–343.4
Tracheophyta	450.8–431.2	450.8–430.4
Lycopodiophyta	432.5–392.8	431.2–392.8
Euphyllophyta	437.6–402.2	435.7–402.2
Monilophyta	411.5–384.9	409.3–384.9
Spermatophyta	365.0–330.9	365.0–329.8
Acrogymnospermae	337.2–308.4	335.9–308.4
Pinopsida	301.3–172.4	302.8–172.1
Angiospermae	246.5–197.5	246.6–195.4
Mesangiospermae	180.4–139.5	177.6–139.2
Magnoliids	149.9–118.9	149.1–119.1
Piperales	103.7–51.4	106.7–50.6
Eudicotyledoneae	125.0–119.7	124.2–119.7
Monocotyledoneae	128.5–114.5	128.5–114.6

N/A, not applicable.

Cambrian to Early Ordovician interval and, regardless of topology, all four major lineages of land plants had diverged by the late Silurian. These dates are older than those used in the latest biogeochemical models (6, 8), and thus our results have implications for simulations of atmospheric chemistry and climate during the Paleozoic.

Results

Topology. The competing hypotheses of relationships among bryophytes and tracheophytes all produce congruent age estimates across the phylogeny (Fig. 2 and Tables 2 and 3). Age estimates of key nodes (Embryophyta, Tracheophyta) are very similar regardless of the underlying topology (Fig. 2 and Tables 2 and 3). At the full range of uncertainty across topologies, the 95% highest posterior density (HPD) of ages for the embryophyte node ranges from the mid-Cambrian (Series 2; 515.2 Ma) to Early Ordovician (473.5 Ma) (Table 2), with the bulk of the distributions in the Cambrian (Fig. 2). There is a slightly higher variance in the estimated age of tracheophytes between the different topologies, but there is overlap in all of the 95% HPD age ranges (Fig. 2 and Tables 2 and 3). Estimates for the age of crown tracheophytes range from Late Ordovician (Katian; 450.8 Ma) to the latest Silurian (419.3 Ma).

The two main hypotheses of early land plant relationships (monophyletic bryophytes and hornworts-sister) give congruent estimates for all nodes across the tree (Fig. 3 and Table 3). For example, the age estimates based on the two topologies are similar for Viridiplantae (972.4 Ma to 669.9 Ma), Streptophyta (890.9 Ma to 629.1 Ma), and Angiospermae (246.6 Ma to 195.4 Ma).

Dataset Size. Infinite site plots describe the relationship between clade age and uncertainty (95% HPD of clade age estimates). As the volume of sequence data increases, it is anticipated that clade age estimates should converge on a straight line, with residual dispersion reflecting uncertainty in calibrations that cannot be overcome by additional sequence data (38). We explored the impact of dataset size based on the monophyletic bryophytes topology, trimming the original dataset (1.7 million nucleotides) based on taxon completeness by 50%, 75%, 99%, and 99.9%. As expected, the resulting infinite sites plots reveal greater uncertainty ($<R^2$) associated with the smallest datasets (Fig. 4) and greatest disparity between the smallest and largest datasets (*SI Appendix*, Fig. S5). However, these differences are small, and, generally, the infinite sites plots indicate that the clade age estimates are effectively insensitive to three orders of difference in the number of nucleotides used in the analysis.

Dating Strategies. Across all alternative dating strategies, the age estimate for crown Embryophyta ranges from 583.1 Ma to 470.0 Ma (Fig. 5 and Table 4), which is larger than the range across the different topologies (515.2 Ma to 473.5 Ma). The greatest variance is seen when the embryophyte constraint is removed, resulting in older age estimates in the hornworts–sister topology, with an age distribution that stretches into the Proterozoic (to the middle Ediacaran), compared with the bulk of the distributions that fall within the Cambrian for all other age estimates (Fig. 5).

We employed different parametric distributions (uniform, Cauchy, skew-t) to express the prior probability of divergence timing relative to the minimum and soft maximum constraints. This often has a dramatic impact on divergence time estimates (39–41); however, different prior distributions have minimal impact on age estimates for embryophytes. The largest difference is seen with the younger age estimates produced using the skew-t distribution (Fig. 5), but both the skew-t and Cauchy models produce younger mean estimates for embryophytes compared with the uniform distribution (Fig. 5). Similarly, there is a younger estimated age for tracheophytes with the skew-t and Cauchy models compared with the uniform distribution (Fig. 5). The age of the tracheophyte node ranges from 472.2 Ma to 422.4 Ma across all alternative dating strategies.

Discussion

Our results demonstrate that divergence time analyses of early land plant evolution are largely insensitive to tree topology and dataset size; however, they show some sensitivity to calibration strategy and, in particular, the calibration on crown Embryophyta. This clearly demonstrates the informative nature of the calibration on crown Embryophyta, which is comparatively narrow in its temporal range (515.5 Ma to 469.0 Ma). The soft maximum constraint on the age of this clade is based on the maximum age of the oldest-possible nonmarine palynomorphs, encompassing all possible total-group embryophyte records (*SI Appendix*). Land plant spores are encountered commonly among marine palynomorph assemblages, and they have the same fossilization and sampling potential as acritarchs. However, the oldest-possible embryophyte records are preceded stratigraphically by thick sequences bearing only marine palynomorphs. These marine palynomorphs demonstrate that the conditions required for preserving embryophyte remains obtained and, thus, the absence of land plant spores constitutes evidence that embryophytes were not present at this time (42). Thus, we discount the results of the divergence time analyses in which the embryophyte calibration is not employed. Similarly, the skew-t and Cauchy distributions, which reflect a nonuniform probability of divergence timing between the minimum and maximum constraints, suggest younger clade ages. However, these nonuniform distributions are unduly informative, since we have no insight or additional evidence that might inform the probability of the time of divergence between minimum and maximum constraints. Hence, we reject the ensuing results in favor of those based on a uniform distribution which reflects equal probability of divergence timing between minimum and maximum constraints. Since the remaining sources of uncertainty have little impact, a holistic timescale encompassing all relevant uncertainties is, effectively, that represented in Fig. 2. It is difficult to foresee how higher precision can be achieved while also maintaining accuracy. We have shown that additional sequence data and

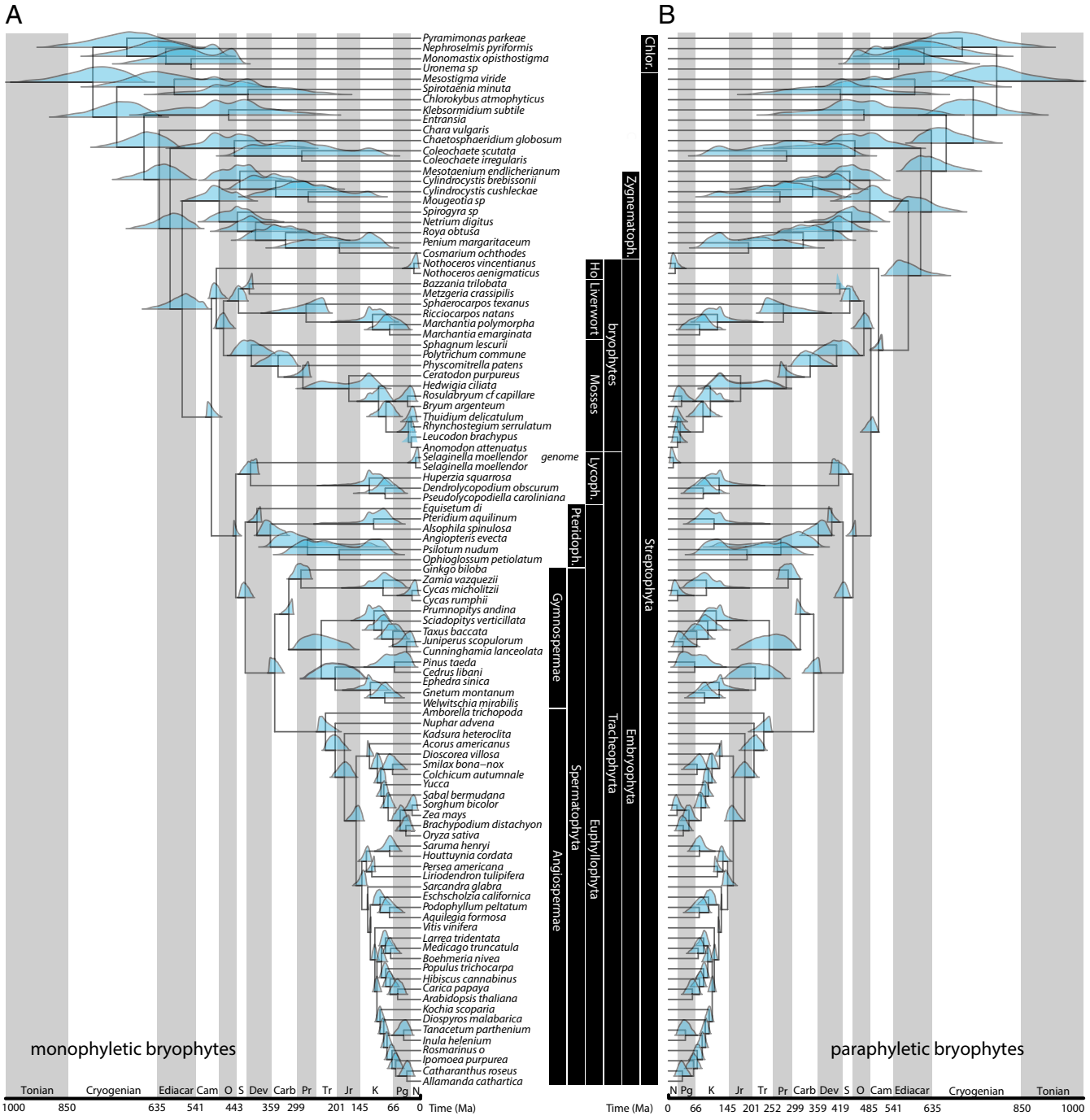


Fig. 3. Detailed phylogenies showing the congruent age estimates produced using the monophyletic (*A*) and hornworts–sister (*B*) topologies.

topological uncertainty have little material impact, both perhaps as a consequence of the short temporal succession of clade divergences among early embryophytes and attendant issues such as incomplete lineage sorting. Improved taxon sampling among liverworts and hornworts (especially) is likely to yield more precise estimates for divergences among bryophytes on some topologies, as would improved sampling of their fossil record—which our analyses predict to extend deep into the Lower Paleozoic.

It is possible that a Total Evidence approach (43), integrating living and fossil species, both morphological and molecular data and evolutionary models, will leverage some increased precision. Perhaps more importantly, such an approach might provide a means of more precisely dating the origin of land plant body plan innovations (e.g., stomata, leaves, rooting systems) that have been considered influential in the evolution of the Earth System (44). In the interim, our evolutionary timescale achieves precision while also integrating all of the principal sources of uncertainty, providing a framework for inferring plant evolutionary history, the veracity of its fossil record, and the impact of phytoterrestrialization on the evolution of global biogeochemical cycles.

The Origin of the Embryophytes and Tracheophytes. Considering the 95% HPDs of divergence times across all topologies, the origin of crown embryophytes is dated to 515.1 Ma to 470.0 Ma (middle Cambrian–Early Ordovician). However, all of the mean estimated

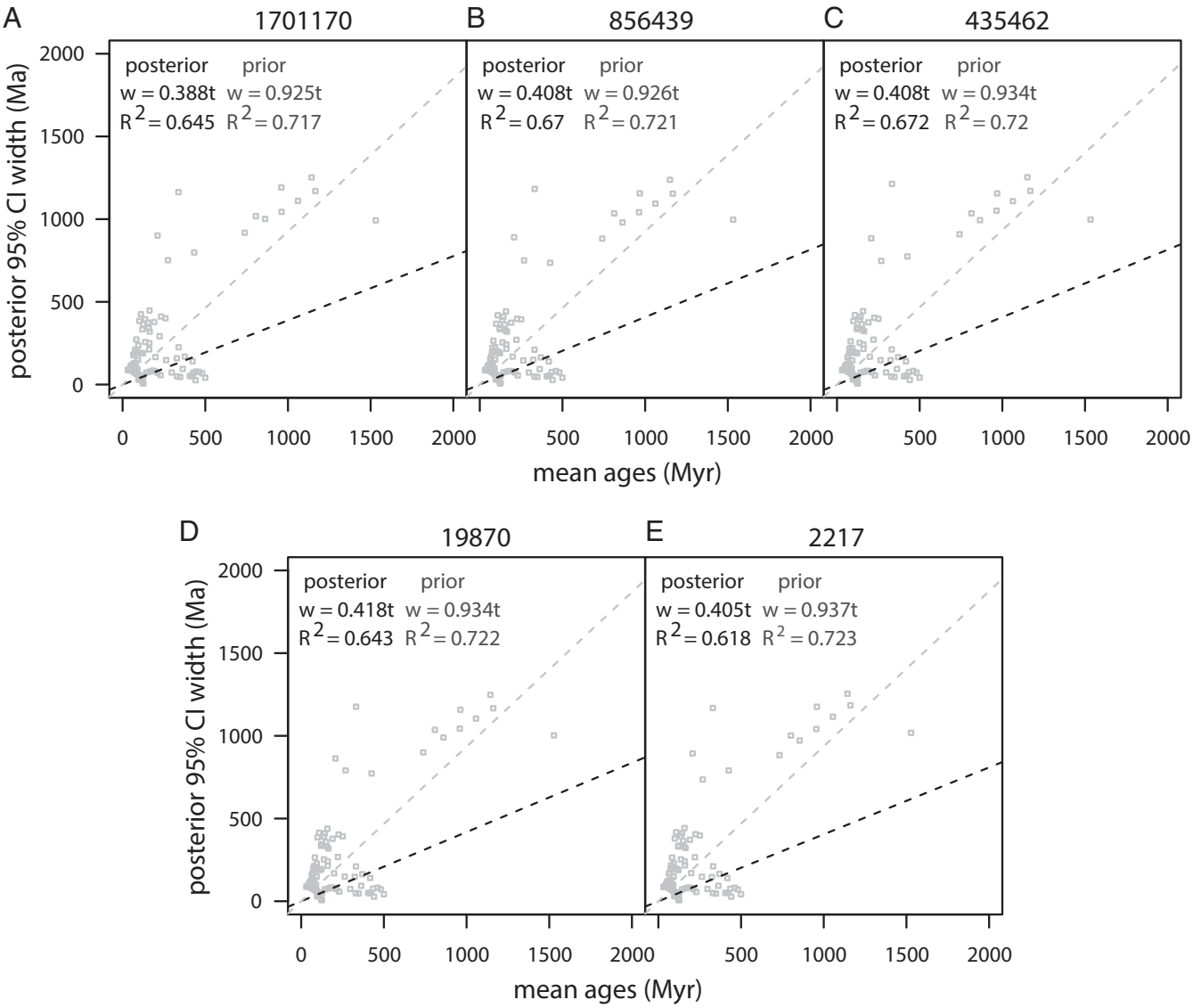


Fig. 4. Infinite site plots showing the effects of including more sequence data on the precision of age estimates. All ages are plotted using the monophyletic bryophytes topology with (*A*) datasets including all sites, and datasets trimmed so sequences are complete for (*B*) 50%, (*C*) 75%, (*D*) 95%, and (*E*) 99.9% of taxa.

ages are resolved within the Phanerozoic across all alternative topologies and dating strategies, and the majority are dated to around 500 Ma (middle Cambrian Series 2). Only one analysis has a 95% HPD that stretches into the Proterozoic. The full span of age estimates for the crown tracheophyte node is 472.2 Ma to 419.3 Ma (Floian, Early Ordovician to the late Silurian). Only one analysis has a 95% HPD that stretches to the Early Ordovician, with those using a uniform prior resulting in estimated mean ages close to the Ordovician–Silurian boundary (~444 Ma). The span of the tracheophyte stem lineage ranges across all analyses from 25.1 My to 60.0 My; these intervals are shorter for the paraphyletic topology than the monophyletic bryophytes topology (35.5 My and 51.6 My, respectively) (*SI Appendix*, Fig. S6).

Impacts of Alternative Topologies and Dating Strategies on Divergence Time Estimates. The impact of analytical uncertainty on the estimated age of Embryophyta is minimized by the use of carefully selected temporal information from the fossil record. Differences in topology had a minimal impact on divergence time estimates for Embryophyta (Fig. 5 and Table 2). For each topology, the posterior age estimates conform largely to the specified calibration constraints on clade age (~511 Ma to 469 Ma). Potential differences in age estimates for embryophytes only appear when the specified age constraint for this node is removed. On the hornworts–sister topology, age estimates for Embryophyta extend into the Proterozoic without the embryophyte calibration, whereas the monophyletic bryophytes topology yields congruent age estimates with or without the user-applied embryophyte age constraint (Fig. 5). Thus, topology can influence the estimated ages for nodes, but only when we ignore germane evidence from the fossil record. Therefore, the use of well-researched and justified fossil constraints, when incorporated alongside tests of model uncertainty, adds confidence in the conclusion of an Early Phanerozoic origin for embryophytes.

There are only minor differences across topologies for the estimated age of tracheophytes, as all trees produce comparable mean estimates (Table 2). One topology, hornworts–liverworts–mosses, produces a younger age from the 95% HPD interval (419 Ma) compared with all other trees (430 Ma), but this

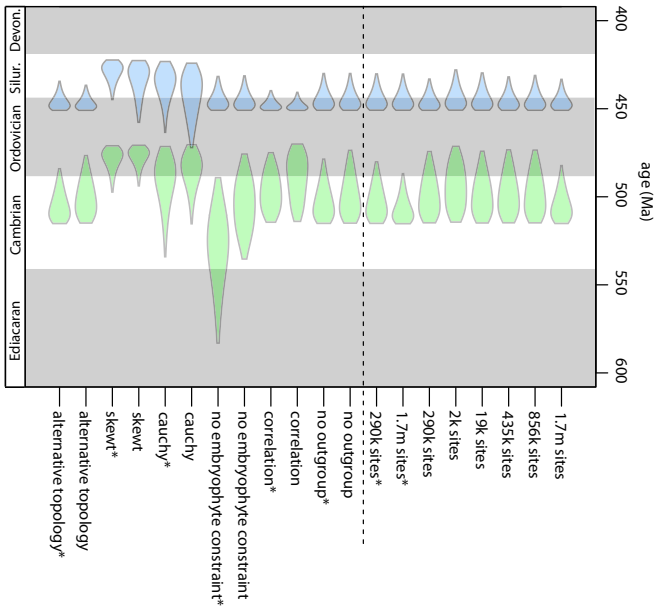


Fig. 5. The estimated ages of embryophyte and tracheophyte divergence is more variable due to differences in modeling compared with differences in dataset size or topology. Using the monophyletic topology, the impact on age estimation was tested by using alternative strategies to model substitution rates, age constraints, and by excluding outgroups. An asterisk (*) denotes analysis performed on hornworts–sister topology.

younger age is anomalous (i.e., slightly younger than the minimum derived directly from fossil evidence at 420.7 Ma) and has little overall support; the bulk of the posterior age of tracheophytes for the hornworts–liverworts–mosses tree is above 430 Ma.

Comparisons with the Fossil Record. The first unequivocal embryophyte body fossil taxon, *Cooksonia* cf. *pertoni*, appears in the Wenlock [minimum age of 426.9 Ma (45)]. The first account of crown tracheophyte body fossils is shortly after, in the Ludlow [minimum age of 420.7 Ma (46)], followed by an apparent explosion of diversity in the Early Devonian (13). Our mean age estimates are older for both nodes, by 40 My for the embryophytes and 20 My for crown tracheophytes. However, in both cases, this is a consequence of a dearth of continental lithofacies before the late Silurian–Early Devonian (47). The earliest known fossils of embryophyte affinity are permanently fused tetrahedral tetrad cryptospores [*sensu stricto* Steemans (48), Wellman (49)] that have a long history of occurrences within marine deposits (13) from the Middle Ordovician [Dapingian; 469 Ma (50)]. Cryptospores of unclear affinity from the Cambrian [*sensu stricto* Strother (51)], while not considered unequivocally embryophyte, informed our soft maximum constraint (515.5 Ma). Our middle Cambrian–Early Ordovician estimate for the origin of crown embryophytes is compatible with an embryophyte interpretation; however, our results do not suggest that they reflect a protracted cryptic earlier evolutionary history. Likewise, the dispersed record of trilete spores that first appear in the Katian (Late Ordovician) (52), followed by an explosion of diversity in the Silurian (13), indicates an earlier origin for tracheophytes that is congruent with our estimates.

The main challenge in testing our divergence time estimates for the bryophyte lineages is their very poor representation in the rock record (13). Nevertheless, our results establish a predictive temporal framework for the stratigraphic intervals in which to prospect for fossils implied by the ghost lineages in our evolutionary timescale. Regardless of the topology, we date the first and second divergences within the bryophytes between 496.5 Ma and 456.2 Ma (late Cambrian–Late Ordovician) and 478.7 Ma and 438.0 Ma (Early Ordovician–early Silurian), respectively. The oldest credible candidate bryophyte fossil is the Pragian (Early Devonian) *Riccardiothallus devonicus* (53), although the security of its classification is limited by preservation of only gross morphology. The mismatch between the estimated ages and unequivocal fossil finds is contributed to by their low fossilization potential, principally because bryophytes do not biosynthesize lignin. When body fossils occur, they are often too poorly preserved to allow recognition of synapomorphies. However, some extant bryophytes produce permanent tetrads and dyads (54, 55) similar to the cryptospores. The wall ultrastructure of cryptospores, known from as early as the Middle Ordovician, is similar to the multilaminate walls observed in permanent tetrads produced by extant liverworts, such as *Sphaerocarpos* (56). The presence of liverwort-like spores in the Middle Ordovician is not incongruent with the estimated dates of divergence of the liverworts across all topologies in our analyses. Sporangia described from the Late Ordovician of Oman are significant fragments of plant anatomy recovered from very rare instances of nonmarine Ordovician rocks (57). The spore masses contain either dyads or tetrads, the former displaying multilaminate walls, and most specimens preserve at least a partial covering, making it very difficult to argue that they are anything but land plant sporangia (57, 58). Unfortunately, our understanding of

Table 4. 95% HPD age estimates for embryophytes and tracheophytes in analyses after removing all nonembryophyte lineages, employing a correlated clock model, and applying different strategies for the shape of prior node age constraints (uniform unless stated)

Dating strategies	Embryophytes, Ma	Tracheophytes, Ma
Monophyletic no outgroup	515.0–473.6	450.8–430.1
Hornworts–sister no outgroup	515.1–478.6	450.8–430.1
Monophyletic correlation	514.0–470.0	450.9–440.7
Hornworts–sister correlation	514.4–475.0	450.9–439.8
Monophyletic no embryophyte constraint	535.3–475.7	450.8–431.4
Hornworts–sister no embryophyte constraint	583.1–489.2	450.8–431.7
Monophyletic cauchy	515.3–470.4	472.2–424.2
Hornworts–sister cauchy	534.0–471.4	463.4–423.2
Monophyletic skew-t	493.8–470.7	457.7–422.7
Hornworts–sister skew-t	497.3–471.1	444.8–422.4
Monophyletic (*Chara*–embryophytes)	514.9–476.6	450.9–436.7
Hornworts–sister (*Chara*–embryophytes)	515.2–484.1	450.9–434.5

There is greater variance when these uncertainties are used compared with the smaller variance seen on dating analyses using the alternative topologies.

the parent plants of cryptospores, the cryptophytes, is restricted to much later charcoal Lagerstätten in the Pridoli and Lochkovian (59, 60). These fossil plants possess a combination of both bryophytic and tracheophytic characters, and thus their taxonomic position is currently unclear (60). The confirmation of the main synapomorphy for the tracheophytes, the presence of vascular tissues, is particularly difficult to demonstrate, due to the minute size and fused nature of these fossils.

Theories on the process of terrestrialization have long argued for a close temporal relationship between the emergence of land plants and terrestrial animals, particularly arthropods, substantiated by their approximately concurrent first fossil occurrence in terrestrial facies (61, 62). However, this is likely an instance of pseudocongruence, with lineages of differing antiquity exhibiting coeval stratigraphic first occurrences because of secular variation in the preservation of Lower Paleozoic terrestrial facies (40). Thus, a shift from dominantly marine to terrestrial facies results in a telescoped first stratigraphic appearance of disparate terrestrial lineages (63). The results of our divergence time analyses indicate a much earlier (~70 My to 80 My) origin of land plants, but, surprisingly, this remains congruent with the latest divergence time estimates for three or four independent transitions to terrestrialization among arthropod lineages (hexapods, arachnids, and, perhaps, twice among myriapods) (64). Thus, although our results corroborate the view that the early fossil records of terrestrial arthropods and land plants are temporally misleading, they also corroborate the hypothesis of a close temporal relationship between the emergence of land plants and terrestrial animals, with plants creating habitats suitable for terrestrial arthropods.

Comparisons with Previous Studies. Previous analyses indicate either a Proterozoic (mainly Cryogenian) (65–67) or Phanerozoic (68–70) origin of the embryophytes. Of the latter, dates range from the Early Ordovician [~474 Ma to 477 Ma (69, 70)] to early Silurian [435 Ma to 425 Ma (68)]. The majority of our results are congruent with a Phanerozoic origin, but with older estimated ages (middle Cambrian; Fig. 5), reflecting the use of Cambrian cryptospores as a soft maximum constraint on crown embryophyte divergence. In comparison with the fossil record, a Phanerozoic origin of the embryophytes is more tenable than the Proterozoic, which is effectively precluded by the absence of embryophyte remains in marine sequences that nevertheless preserve sporopollenin acritarchs (42).

The origin of the crown tracheophytes has been fixed as a calibration point in most previous studies. Estimated ages include the Late Ordovician (446 Ma) (67), mid-Silurian (432 Ma to 434 Ma) (69), and late Silurian (423.95 Ma) (70). Our analyses are most congruent with the older ages estimated by Clarke et al. (67), around the Ordovician–Silurian boundary (Fig. 5), as a result of the application of an older taxon for the calibration [e.g., *Zosterophyllum* instead of *Leclercqia* as in Smith et al. (69)], and a soft maximum age constraint using the first occurrence of trilete spores in the Katian.

Few molecular clock studies focus on bryophyte divergence and, as such, often have restricted analyses to stomatophytes (mosses vs. vascular plants) (65, 66), including very few taxa. Estimates for the first bryophyte divergence begin as early as the Cryogenian (65, 66), with further studies suggesting the Ediacaran to late Cambrian (632 Ma to 499 Ma) (67), late Cambrian to late Silurian (490 Ma to 425 Ma) (68), Late Ordovician (458 Ma) (70), and mid-Devonian (383 Ma) (69). Our age estimates are most congruent with an Early Paleozoic divergence. Where previous studies have included all bryophyte lineages, the second divergence has been estimated from the Early Cambrian–Middle Ordovician (67) (524 Ma to 460 Ma), early Silurian (70) (440 Ma), and Mississippian (69) (335 Ma). Our age estimates are more congruent with the older estimates.

Implications for Hypotheses on the Coevolution of Land Plants and Climate. The evolution and geographical spread of the embryophytes across Paleozoic continents undoubtedly had a major impact upon global biogeochemical cycles. To test hypotheses on the coevolution of land plants and Earth's System, biogeochemical models rely on a well-substantiated phylogeny and timeline of embryophyte divergence and character acquisition. The GEOCARB (3, 4) and COPSE (5) biogeochemical models include parameters for the evolution and geographical spread of tracheophytes and their enhancement of silicate weathering rates, resulting in simulations that show a significant decrease in atmospheric CO_2 levels in the Devonian [from ~16× to ~3× present atmospheric level (PAL)] and the rise in O_2 levels to 1.5× PAL by the end of the Carboniferous. However, these models are undermined by their use of the body fossil record to establish a timescale for plant evolution and innovations. These weaknesses can be overcome by considering the divergence time estimates of key innovations from molecular clock studies.

Our results demonstrate that embryophytes were present on land from the middle Cambrian–Early Ordovician interval, and minimally, by the early Silurian, the four major lineages of land plants had already diverged and were constituents of early cryptogamic ground covers (71). Plants had already evolved key adaptations for survival and proliferation on dry land by the early Silurian (e.g., development of an embryo, alternation of generations, aerial sporophytes, sporophyte branching, cuticle, stomata, vascular tissue, sporopollenin-coated spores), including interactions with early soils and nutrient extraction from minerals (rhizoids, rhizomes, and symbiosis with mycorrhizal fungal partners). The results of our analyses suggest that the majority of these characters had evolved within a middle Cambrian–Early Ordovician interval. Modern cryptogamic covers, that comprise bryophytes, lichens, fungi, algae, and cyanobacteria, are capable of significant mineral weathering (72, 73), in particular via symbiotic mycorrhizal fungal partners accessing phosphorous (7), a limiting nutrient, which results in a positive feedback mechanism with increasing biomass of the host plant. As such, the timing of divergence and weathering capabilities of these early ground covers has been underestimated in these biogeochemical models.

Conclusions

The origin and evolution of land plants has transformed the terrestrial biosphere. Our understanding of the timing and nature of this formative episode is undermined by uncertainties associated with the incompleteness of the plant fossil record and the evolutionary relationships of the living land plant lineages. We establish an evolutionary timescale that integrates over these uncertainties, estimating the living clade of land plants to have emerged in the middle Cambrian–Early Ordovician, and the living clade of vascular plants to have appeared in a Late Ordovician–Silurian interval. These are in close accord with estimates for the timing of terrestrialization of arthropod lineages. These results underscore the importance of taking an integrative approach to the establishment of evolutionary timescales, which can only be derived through application of molecular clock methodology (74). Future attempts to explore the role of plant phylogeny in the evolution of global biogeochemical cycles must integrate this recalibrated timescale for plant evolution, rather than relying on the fossil record alone.

Methods

Dating Analyses. We conducted all dating analyses in MCMCTree within the software PAML version 4.8 (75), and all analyses were prepared using MCMCTreeR in R (https://github.com/PuttickMacroevolution/MCMCTreeR).

Genetic Data. We used two datasets from the published nucleotide alignments of Wickett et al. (14) for all analyses. For the first dataset, we used

the full 852-gene alignment of 1,701,170 nucleotides. We used a subset of these data that were filtered by Wickett et al. (14) to maximize coverage of sites and genes, remove potential contamination, and exclude the third codon position. These data consist of 290,718 nucleotides. Unless specified, all subsequent analyses were conducted using the dataset of 1,701,170 nucleotides.

Topology. We estimated topology using topological constraints to enforce each of the seven hypotheses (Fig. 1) but leaving all other relationships unconstrained, using the focal dataset of Wickett et al. (14) (290,718 nucleotides, trimmed and third codon removed). For each hypothesis, we constrained tracheophytes, each bryophyte group (liverworts, hornworts, mosses), and the non-embryophytes. With each of these constraints, we left all other relationships as polytomies. We estimated these topologies in RAxML 8.2 (76) in a nonpartitioned, nucleotide GTR + Γ model.

Dataset Size. We explored the impacts of dataset size (number of nucleotides) and site completeness. Plots of infinite sites were used to gauge the potential increase in precision gained by adding more sequence data. We compared infinite site plots of the original sequence data (852 genes, 1.7 million nucleotides) to data we trimmed by site completeness so that only sites complete for 50%, 75%, 99%, and 99.9% of species were included; this produced datasets with 850,000, 435,000, 19,000, and 2,000 nucleotides, respectively. These initial analyses indicated that there is not much effect in adding more sequence data (Fig. 4), and thus, for comparisons of all seven hypotheses, we employed the dataset trimmed by 50% completeness (850,000 nucleotides).

Rate Priors. To incorporate deviations of a strict molecular clock, we set the IGR model that treats branch rates as being samples from independent and identically distributed log-normal distributions (77, 78). This distribution is given a prior mean rate for branches (μ), and variance σ^2 that models the overall rate variability on branches across the phylogeny. In MCMCTree, the mean rate is given a prior gamma distribution with user-specified shape and scale. To obtain a suitable prior on the substitution rate (μ), we compared the pairwise distance between *Arabidopsis thaliana* and *Rhynchostegium serrulatum* using the GTR + Γ + F model in baseml version 4.8 (75). For the smaller dataset, this resulted in a substitution rate of 0.08^{-10} changes per nucleotide site per year after assuming a divergence time of 469 Ma. In the larger dataset, this value was 0.09^{-10} nucleotide substitutions per site per year. As in dos Reis et al. (79), we fixed the shape parameter of the gamma distribution prior on rate to 2, and, from this, set the scale parameter to 25. For the larger dataset, these figures were set to shape 2 and scale 22. We estimated these parameters for each of the subsets of the larger dataset. We set the prior on rate variability (σ^2) as a gamma distribution with shape 1 and scale 10.

Time Priors. For the priors on branching times, we set the prior birth–death process with parameters of birth = 1, death = 1, and fraction of sampled species = 0, which produce a uniform kernel for the branching times. The time prior or the prior for divergence times for all nodes in the tree is generated in conjunction with the specified node age densities based on the fossil record. The specified calibration densities and the effective time prior can be very different (41). To ensure our priors on divergence times were appropriate, we ran the model without sequence data to obtain the effective priors.

Fossil Ages and Prior Node Distributions. In each analysis (unless stated), we applied temporal node constraints to 37 nodes, including the root. The location of the 37 nodes is shown in *SI Appendix*, Fig. S1. We applied node distributions using minimum and maximum constraints following protocols outlined in Parham et al. (37). For full phylogenetic and age justifications of each fossil calibration, see *SI Appendix*, *SI Methods* and Tables S1–S7.

Three strategies were applied to specifying the prior distributions on node ages. In strategy *i*, uniform distributions were applied to all internal prior node ages with a hard minimum age and a soft maximum age, allowing 0.001% probability of an age younger or older than the given minima and maxima. For strategy *ii*, we applied skew-normal distributions with the mode of the distribution above the minimum age and 0.001% and 97.5% probability tails at the maximum and minimum ages. For strategy *iii*, we applied Cauchy distributions with a hard minimum and a 97.5% probability at the maximum age. For strategy *iii*, the root node was set as a uniform distribution. For each strategy, we assessed the shape of prior and posterior distributions on the 37 nodes to which we applied data from the fossil record (shown for the hornworts–sister topology, *SI Appendix*, Figs. S2–S4). The specific parameters used for input into MCMCTree are shown in *SI Appendix*, Tables S8 and S9.

Analyses of large datasets can be highly time-consuming. Therefore, we implemented the approximate likelihood calculations available in MCMCTree (80, 81). We obtained estimates of branch lengths in baseml (82), and, in the program, these maximum likelihood estimates are then used to obtain the gradient and Hessian matrix of the branch lengths. These estimates were then used to calculate the approximate likelihood (81) in the divergence time analyses.

Dating Strategies. Two of the key nodes we were primarily interested in dating were crown embryophytes and crown tracheophytes. We conducted several sensitivity analyses to explore any potential variation in the age estimates. We tested the effect of removing the nonembryophyte (algal) species from the analysis so the embryophyte node became the root node. In a separate analysis, we removed the user-applied node constraint for embryophytes. We also explored the impact of applying a correlated clock model to the data (80). Additionally, we explored the effect of using topologies based on the maximum likelihood tree search of the 1.7-million nucleotide dataset; the largest difference in this topology is that *Chara vulgaris* is sister to embryophytes rather than Zygnematophyceae in the main analyses. Finally, we explored the effects of codon partition by comparing the posterior age estimates of a single partition (all codons in a single alignment) and a partition of each codon (three alignments for positions 1, 2, and 3). These analyses indicated that partition did not have any meaningful influence on posterior age estimates for all nodes (*SI Appendix*, Fig. S7).

ACKNOWLEDGMENTS. We acknowledge funding from the Natural Environment Research Council Grants NE/N003438/1 and NE/J012610/1, the Biotechnology and Biological Sciences Research Council Grant BB/N000919/1, and Royal Society Wolfson Merit Award (to P.C.J.D.). Z.Y. was supported in part by the Radclie Institute for Advanced Study at Harvard University.

1. Algeo TJ, Scheckler SE (1998) Terrestrial-marine teleconnections in the Devonian: Links between the evolution of land plants, weathering processes, and marine anoxic events. *Philos Trans R Soc Lond B Biol Sci* 353:113–128.
2. Morris JL, et al. (2015) Investigating Devonian trees as geo-engineers of past climates: Linking palaeosols to palaeobotany and experimental geobiology. *Palaeontology* 58: 787–801.
3. Berner RA, Kothavala Z (2001) GEOCARB III: A revised model of atmospheric CO_2 over Phanerozoic time. *Am J Sci* 301:182–204.
4. Berner RA (2006) GEOCARBSULF: A combined model for Phanerozoic atmospheric O_2 and CO_2. *Geochim Cosmochim Acta* 70:5653–5664.
5. Bergman NM, Lenton TM, Watson AJ (2004) COPSE: A new model of biogeochemical cycling over Phanerozoic time. *Am J Sci* 304:397–437.
6. Lenton TM, Crouch M, Johnson M, Pires N, Dolan L (2012) First plants cooled the Ordovician. *Nat Geosci* 5:86–89.
7. Quirk J, et al. (2015) Constraining the role of early land plants in Palaeozoic weathering and global cooling. *Proc Biol Sci* 282:20151115.
8. Lenton TM, et al. (2016) Earliest land plants created modern levels of atmospheric oxygen. *Proc Natl Acad Sci USA* 113:9704–9709.
9. Labandeira CC (2013) A paleobiologic perspective on plant-insect interactions. *Curr Opin Plant Biol* 16:414–421.
10. Selosse M-A, Strullu-Derrien C, Martin FM, Kamoun S, Kenrick P (2015) Plants, fungi and oomycetes: A 400-million year affair that shapes the biosphere. *New Phytol* 206: 501–506.
11. Retallack GJ (2003) Soils and global change in the carbon cycle over geological time. *Treatise Geochem* 5:1–28.
12. Gibling MR, Davies NS (2012) Palaeozoic landscapes shaped by plant evolution. *Nat Geosci* 5:99–105.
13. Kenrick P, Wellman CH, Schneider H, Edgecombe GD (2012) A timeline for terrestrialization: Consequences for the carbon cycle in the Palaeozoic. *Philos Trans R Soc Lond B Biol Sci* 367:519–536.
14. Wickett NJ, et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA* 111:E4859–E4868.
15. Mishler BD, Churchill SP (1984) A cladistic approach to the phylogeny of the 'bryophytes.' *Brittonia* 36:406–424.
16. Garbary DJ, Renzaglia KS (1988) *Bryology for the Twenty-First Century*, eds Bates JW, Ashton NW, Duckett JG (Maney, Leeds, UK), pp 45–63.
17. Renzaglia KS, Nickrent DL, Garbary DJ, Garbary DJ; Duff RJT (2000) Vegetative and reproductive innovations of early land plants: Implications for a unified phylogeny. *Philos Trans R Soc Lond B Biol Sci* 355:769–793.
18. Qiu YL, Cho Y, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394:671–674.
19. Mishler BD, et al. (1994) Phylogenetic relationships of the 'green algae' and 'bryophytes.' *Ann Mo Bot Gard* 81:451–483.
20. Mishler BD, et al. (1992) A molecular approach to the phylogeny of bryophytes: Cladistic analysis of chloroplast-encoded 16S and 23S ribosomal RNA genes. *Bryologist* 95:172–180.

21. Karol KG, McCourt RM, Cimino MT, Delwiche CF (2001) The closest living relatives of land plants. *Science* 294:2351–2353.
22. Lewis LA, Mishler BD, Vilgalys R (1997) Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbc*L. *Mol Phylogenet Evol* 7:377–393.
23. Qiu Y-L, et al. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* 103:15511–15516.
24. Qiu YL, et al. (2007) A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *Int J Plant Sci* 168: 691–708.
25. Chang Y, Graham SW (2011) Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *Am J Bot* 98:839–849.
26. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol Biol* 14:23.
27. Liu Y, Cox CJ, Wang W, Goffinet B (2014) Mitochondrial phylogenomics of early land plants: Mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Syst Biol* 63:862–878.
28. Waters DA, Buchheim MA, Dewey RA, Chapman RL (1992) Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. *Am J Bot* 79:459–466.
29. Zhong B, et al. (2014) Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Mol Biol Evol* 31: 177–183.
30. Hedderson TAJ, Chapman RL, Rootes WL (1996) Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. *Plant Syst Evol* 200: 213–224.
31. Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J* 15:1403–1411.
32. Nishiyama T, Kato M (1999) Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. *Mol Biol Evol* 16:1027–1036.
33. Beckert S, Steinhauser S, Muhle H, Knoop V (1999) A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial*nad*5 gene. *Plant Syst Evol* 218:179–192.
34. Nickrent DL, Parkinson CL, Palmer JD, Duff RJ (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol Biol Evol* 17:1885–1895.
35. Nishiyama T, et al. (2004) Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol Biol Evol* 21:1813–1819.
36. Cox CJ, Li B, Foster PG, Embley TM, Civán P (2014) Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Syst Biol* 63:272–279.
37. Parham JF, et al. (2012) Best practices for justifying fossil calibrations. *Syst Biol* 61: 346–359.
38. Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 23: 212–226.
39. Warnock RCM, Yang Z, Donoghue PCJ (2012) Exploring uncertainty in the calibration of the molecular clock. *Biol Lett* 8:156–159.
40. Inoue J, Donoghue PCJ, Yang Z (2010) The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst Biol* 59:74–89.
41. Warnock RC, Parham JF, Joyce WG, Lyson TR, Donoghue PC (2015) Calibration uncertainty in molecular dating analyses: There is no substitute for the prior evaluation of time priors. *Proc Biol Sci* 282:20141013.
42. Gray J, Boucot AJ (1978) The advent of land plant life. *Geology* 6:489–492.
43. Ronquist F, et al. (2012) A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst Biol* 61:973–999.
44. Berner RA (1997) The rise of plants and their effect on weathering and atmospheric CO_2. *Science* 276:544–546.
45. Edwards D, Feehan J (1980) Records of *Cooksonia*-type sporangia from late Wenlock strata in Ireland. *Nature* 287:41–42.
46. Kotyk ME, Basinger JF, Gensel PG, de Freitas TA (2002) Morphologically complex plant macrofossils from the Late Silurian of Arctic Canada. *Am J Bot* 89:1004–1013.
47. Smith AB, McGowan AJ (2007) The shape of the Phanerozoic marine palaeodiversity curve: How much can be predicted from the sedimentary rock record of Western Europe? *Palaeontology* 50:765–774.
48. Steemans P (2000) Miospore evolution from the Ordovician to the Silurian. *Rev Palaeobot Palynol* 113:189–196.
49. Wellman CH (2010) The invasion of the land by plants: When and where? *New Phytol* 188:306–309.
50. Rubinstein CV, Gerrienne P, de la Puente GS, Astini RA, Steemans P (2010) Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). *New Phytol* 188:365–369.
51. Strother PK (2016) Systematics and evolutionary significance of some new cryptospores from the Cambrian of eastern Tennessee, USA. *Rev Palaeobot Palynol* 227: 28–41.
52. Steemans P, et al. (2009) Origin and radiation of the earliest vascular land plants. *Science* 324:353.
53. Guo C-Q, et al. (2012) *Riccardiothallus devonicus* gen. et sp. nov., the earliest simple thalloid liverwort from the lower Devonian of Yunnan, China. *Rev Palaeobot Palynol* 176–177:35–40.
54. Renzaglia KS, et al. (2015) Permanent spore dyads are not a 'thing of the past'; on their occurrence in the liverwort *Haplomitrium* (Haplomitriopsida). *Bot J Linn Soc* 179: 658–669.
55. Renzaglia KS, Lopez RA, Johnson EE (2015) Callose is integral to the development of permanent tetrads in the liverwort *Sphaerocarpos*. *Planta* 241:615–627.
56. Taylor WA (1995) Ultrastructure of *Tetrahedraletes medinensis* (Strother and Traverse) Wellman and Richardson, from the upper Ordovician of southern Ohio. *Rev Palaeobot Palynol* 85:183–187.
57. Wellman CH, Osterloff PL, Mohiuddin U (2003) Fragments of the earliest land plants. *Nature* 425:282–285.
58. Wellman CH (2003) *Dating the Origin of Land Plants. Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*, Systematics Association Special Volume (Cambridge Univ Press, Cambridge, UK), Vol 66, pp 119–141.
59. Wellman CH, Edwards D, Axe L (1998) Permanent dyads in sporangia and spore masses from the Lower Devonian of the Welsh Borderland. *Bot J Linn Soc* 127: 117–147.
60. Edwards D, Morris JL, Richardson JB, Kenrick P (2014) Cryptospores and cryptophytes reveal hidden diversity in early land floras. *New Phytol* 202:50–78.
61. Edwards D, Selden PA (1993) The development of early terrestrial ecosystems. *Bot J Scotl* 46:337–366.
62. Edwards D, Selden PA, Richardson JB, Axe L (1995) Coprolites as evidence for plant-animal interaction in Siluro-Devonian terrestrial ecosystems. *Nature* 377:329–331.
63. Holland SM (2016) The non-uniformity of fossil preservation. *Philos Trans R Soc Lond B Biol Sci* 371:20150130.
64. Lozano-Fernandez J, et al. (2016) A molecular palaeobiological exploration of arthropod terrestrialization. *Philos Trans R Soc Lond B Biol Sci* 371:20150133.
65. Heckman DS, et al. (2001) Molecular evidence for the early colonization of land by fungi and plants. *Science* 293:1129–1133.
66. Hedges SB, Blair JE, Venturi ML, Shoe JL (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol* 4:2.
67. Clarke JT, Warnock RC, Donoghue PC (2011) Establishing a time-scale for plant evolution. *New Phytol* 192:266–301.
68. Sanderson MJ (2003) Molecular data from 27 proteins do not support a Precambrian origin of land plants. *Am J Bot* 90:954–956.
69. Smith SA, Beaulieu JM, Donoghue MJ (2010) An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc Natl Acad Sci USA* 107:5897–5902.
70. Magallón S, Hilu KW, Quandt D (2013) Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am J Bot* 100:556–573.
71. Edwards D, Cherns L, Raven JA, Smith A (2015) Could land-based early photosynthesizing ecosystems have bioengineered the planet in mid-Palaeozoic times? *Palaeontology* 58:803–837.
72. Elbert W, et al. (2012) Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nat Geosci* 5:459–462.
73. Porada P, Weber B, Elbert W, Pöschl U, Kleidon A (2014) Estimating impacts of lichens and bryophytes on global biogeochemical cycles. *Global Biogeochem Cycles* 28:71–85.
74. Donoghue PC, Yang Z (2016) The evolution of methods for establishing evolutionary timescales. *Philos Trans R Soc Lond B Biol Sci* 371:20160020.
75. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
76. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
77. Rannala B, Yang Z (2007) Inferring speciation times under an episodic molecular clock. *Syst Biol* 56:453–466.
78. Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Mol Biol Evol* 24:2669–2680.
79. dos Reis M, et al. (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr Biol* 25:2939–2950.
80. Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15:1647–1657.
81. dos Reis M, Yang Z (2011) Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol* 28:2161–2172.
82. Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* 39:306–314.

Trends in Plant Science

CellPress REVIEWS

Review

Whole-Genome Duplication and Plant Macroevolution

James W. Clark[1] and Philip C.J. Donoghue[1,*]

Whole-genome duplication (WGD) is characteristic of almost all fundamental lineages of land plants. Unfortunately, the timings of WGD events are loosely constrained and hypotheses of evolutionary consequence are poorly formulated, making them difficult to test. Using examples from across the plant kingdom, we show that estimates of timing can be improved through the application of molecular clock methodology to multigene datasets. Further, we show that phenotypic change can be quantified in morphospaces and that relative phenotypic disparity can be compared in the light of WGD. Together, these approaches facilitate tests of hypotheses on the role of WGD in plant evolution, underscoring the potential of plants as a model system for investigating the role WGD in macroevolution.

Highlights

WGD events continue to be discovered throughout plant systematic diversity as a consequence of genome sequencing programs.

The absolute timing of WGD events remains poorly constrained and poorly understood, but many hypotheses regarding the role of WGD in plant evolution depend on precise estimates.

The role of WGD in facilitating diversification has a strong theoretical basis but remains to be rigorously tested, and WGD in species-poor lineages cannot be ignored.

WGD as a driver of plant morphological diversity is an appealing hypothesis, but requires a framework which can quantify morphological variation between lineages and through time.

Whole-Genome Duplication

WGD encompasses multiple processes that lead to the formation of a polyploid organism with three or more sets of the base chromosome number. It has been invoked as a cause of macroevolutionary change [1], explaining everything from extinction resistance to fundamental evolutionary innovation. WGD has been proposed as a driver of diversity [2,3], herbivore interactions [4], geographic expansions [5], climatic niche shifts [6], and of facilitating lineage longevity [7]. Clustering of WGD events around the Cretaceous–Paleogene (K–Pg) interval has led to the hypothesis that genome duplication may have facilitated evolutionary success in the wake of the mass extinction event at the end of the Cretaceous [8,9] (Box 1). Equally, however, it is possible that the extensive history of WGD in plant evolution is incidental or inconsequential, and there are examples, such as mosses and horsetails [7,10], where a macroevolutionary-scale phenotypic impact is not evident. Ancient WGD events (paleopolyploidy) first appeared to be rare [11], but newly sequenced genomes have revealed duplication in an increasing diversity of plant lineages [12,13]. However, with few exceptions, it appears that most of the hypothesized macroevolutionary outcomes have neither been tested nor formulated as hypotheses that are readily testable, despite the diversity of comparative methods that are available for facilitating such tests. There are multiple emerging models explaining how complexity and novelty may arise through genome duplication [14], although fundamental questions remain as to why the outcomes of WGD are so disparate among lineages and whether the nature of the ploidy event influences the outcome (Box 2). Tests are necessary to quantify the macroevolutionary change in the wake of WGD, or else WGD risks becoming a phenomenon that explains everything and, therefore, nothing.

WGD has occurred across the breadth of eukaryote phylogeny [15–18], but the majority of WGD events have occurred within land plants (Embryophyta) (Figure 1). As such, plants provide very many natural experiments from which it may be possible to develop a general theory on the role of WGD in macroevolution. Patterns of diversification among extant taxa have pointed towards a scenario of rarely successful polyploids [19,20]. However, all members of the most diverse lineage

[1]School of Earth Sciences, University of Bristol, Life Sciences Building, Bristol BS8 1TH, UK

*Correspondence: james.clark@bristol.ac.uk (J.W. Clark) and phil.donoghue@bristol.ac.uk (P.C.J. Donoghue).

Box 1. WGD and the K–Pg Boundary

The distribution of WGD events both across plant phylogeny and through time has revealed in multiple independent lineages that WGD events appear to cluster around the K–Pg boundary (Figure 1). This has led to two related hypotheses: that genome duplication may have conferred an 'extinction resistance' to particular lineages of plants, and that polyploid genomes may have allowed surviving lineages to rise to dominance in the wake of this mass extinction episode.

Polyploid plants are sometimes found towards the edge of species ranges, and polyploid genomes facilitate rapid radiations and invasiveness. Polyploid genomes also possess a 'mutational robustness' relative to diploids, and this may provide short-term advantages which could have allowed them to survive and then thrive. An alternative hypothesis suggests that it is not WGD itself that facilitated extinction resistance, but the coincidence that many newly formed polyploids rely on selfing for reproduction. Selfing is also associated with the extremes of novel habitats, but in the long term is seen as an evolutionary dead end. A return to outbreeding could allow the continued success of these lineages and may also explain the apparent lag between WGD and diversification.

These hypotheses are entirely dependent on the precise timing of each duplication event. As shown in Figure 2, current estimates for the timing of WGD are likely to change given a careful appraisal of the fossil record. As such, until the timing of each WGD event that is considered to lie close to the boundary is re-evaluated, this correlation should be treated with caution.

of land plants, the seed plants (Spermatophyta), are descended from an ancestor that underwent at least one round of WGD [21,22]. Furthermore, within Spermatophyta, another WGD is shared by all flowering plants (angiosperms) [21], and a further WGD is shared in turn by several major clades of flowering plants including the monocots [23], eudicots [24,25], Asteraceae [5,26], Brassicales [27], legumes [28], and the most economically important plants, the grasses [29,30] (Figure 2). The paucity of ancient WGD events that was perceived early in the history of genome sequencing increasingly appears to be an oversight, with denser sampling revealing multiple WGD events during the evolution of taxonomically large and small lineages [6].

Double Dates – The Absolute Timing of WGD

Hypotheses on the role of WGD in plant macroevolution are contingent on the phylogenetic (relative) and geological (absolute) timing of each event. Methods to identify WGD events are

Box 2. The Origins of WGD

Traditionally, polyploids are recognized as originating from a single parent species (autopolyploidy, xx to xxxx) or from two hybridizing species (allopolyploidy, xx + yy to xxyy). Current views maintain that these two outcomes exist along a spectrum, with segmental allopolyploids containing paralogs that display varying levels of synteny [77]. A segmental allopolyploid may form via hybridization between two closely related species, or through the process of homoeologous compensation [77]. Despite potential differences in outcome, both are likely to have had significant effects throughout plant evolution (both processes and their potential evolutionary outcomes have recently been reviewed [97–99]). Based on observations from neopolyploids, there is reason to believe that their outcomes may differ, and it is therefore a priority to establish whether ancient events were a consequence of autopolyploidy or allopolyploidy. Methods to differentiate between the two processes are under development, and in some instances ancient events have been successfully characterized. Genome dominance is a phenomenon observed in allopolyploids, where one subgenome shows lower expression and retention than the other (biased fractionation). Signal of a bias in gene retention between subgenomes could provide evidence for allo- rather than autopolyploidy [100]. Gene-tree methods are also capable of resolving allopolyploid WGDs by considering reticulate patterns of gene-tree evolution [17,101], and in some instances they have been able to identify the most likely parental lineages involved in the hybridization event [102].

The nature of the WGD affects the approach required for dating because auto- and allopolyploidy present different issues. The two subgenomes of an allopolyploid would have diverged at the point of speciation between the two parent lineages, rather than at the hybridization event itself [50,103]. Successful and viable hybrids are more likely to arise between closely related species, giving rise to 'segmental allopolyploids'. However, there are examples of hybridization between distantly related lineages of plants [104], which could lead to a significant overestimation of the age of the WGD. Similarly, as outlined previously, autopolyploidy can lead to a prolonged period of tetrasomic inheritance between ohnologs [59]. In this case there is potential to underestimate the age of the WGD because the ohnologs will only start to diverge once disomic inheritance has occurred, and we date the point at which they diverge rather than the date of duplication.

Glossary

Diploidization: sometimes termed fractionation, this is the period following WGD whereby through rearrangement, silencing, and loss of DNA the genome returns to a diploid expression pattern.
Morphospace: an *n*-dimensional multivariate space describing phenotypes, where points represent taxonomic units and the distances between them their (dis)similarities.
Neofunctionalization: following gene duplication, one copy of the gene takes on a novel function while the other copy continues to perform the previous function.
Paralogs, ohnologs, and homologs: two genes related by descent, typically with similar sequences, are homologs. If they share a 1:1 relationship between species, they are orthologs. If they deviate from this 1:1 relationship as a result of a duplication event, they become paralogs. Paralogs that have derived specifically from a WGD event are termed ohnologs, after Susumu Ohno.
Subfunctionalization: following gene duplication, each duplicate performs part of the original function, and in combination both maintain the original function of the gene.

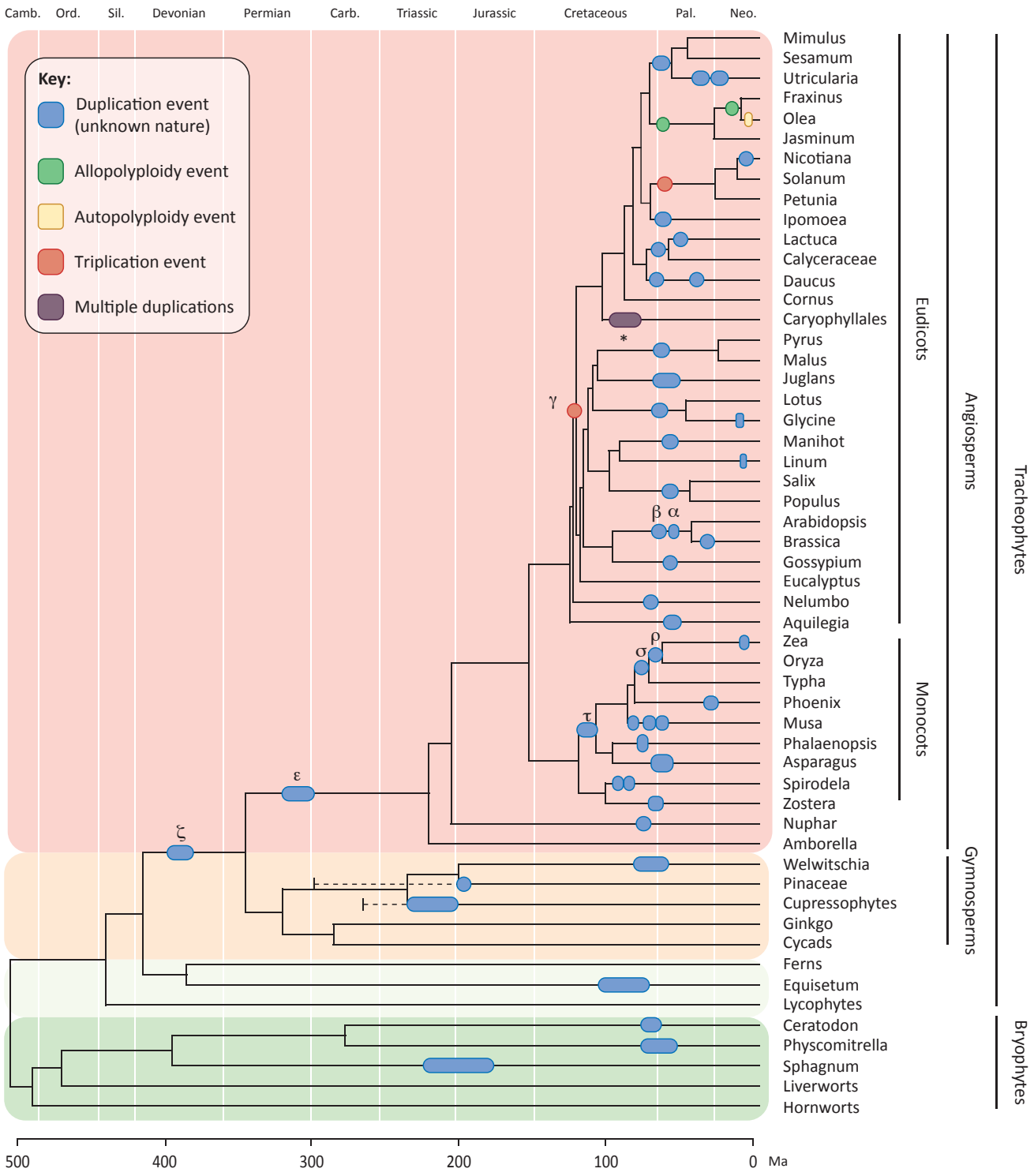


Figure 1. The Distribution of Known Whole-Genome Duplication (WGD) Events within the Plant Kingdom. Most events are shown from Van de Peer *et al.* [91] but have been updated. The length of each bar along the branch indicates the current estimate for its age. Duplication events of unknown origin are shown in navy

(See figure legend on the bottom of the next page.)

many and varied: these include **paralog** (see Glossary) substitution distributions (plots of the synonymous substitution rate, Ks) [31,32], phylogenomics [21], genome size, karyotype, gene copy-number analyses [33,34], and synteny [23,35,36]. Greater sampling of diversity helps resolve the phylogenetic (relative) timing of each WGD, but to refine these hypotheses it is important that their absolute ages are known with accuracy and precision. Absolute ages can be constrained by the age of bracketing speciation events since WGD must have occurred after the divergence of species that have not undergone WGD, and before those living species within the same lineage that have (Figure 2). When taxonomic sampling is dense and the WGD occurred on a short branch (such as with more recent events) this can yield relatively precise age estimates [37]. However, with increasing uncertainty in species divergence time estimates, longer branches, monotypic lineages, or less-dense sampling, it becomes more challenging to directly estimate the timing of a WGD.

As well as being a means to identify and relatively date WGD events, both Ks analyses and phylogenomic methods can be used to directly infer the age of WGD events [32,38–40]. Ks plots represent distributions of rates of synonymous substitutions between paralogs. A peak in the distribution is interpreted as a WGD event, and distributions compared between species can reveal shared duplication events. An external calibration can convert Ks rates into geological time, although this is often done by comparing the position of the peak in Ks rates to ages inferred from phylogenomic dating, for example a Ks value between 0.6 and 1.1 synonymous substitutions per site corresponds to an age of 50–70 Myr. These methods assume a strict rate of molecular evolution, and diff erent rates produce highly variable age estimates. The signature of increasingly ancient WGD events is eroded by sequence saturation, and therefore the dating of more ancient events is prone to error [32]. For example, a WGD event predicted in the early-diverging gymnosperm *Ginkgo biloba* was estimated at between 500 and 700 Ma – pre-dating most estimates for the origin of land plants [41–43].

Phylogenomic approaches exploit the signal of paralogy present in the history of gene families to directly estimate the age of the WGD event [21]. This requires the reconstruction of gene families across multiple species (also termed a phylome [44]) and subjecting them to molecular clock analysis. Molecular clock methodology has typically been applied to dating species divergences but can also be used to date both speciation and duplication events within gene trees. Typically, molecular clock analyses have investigated each gene family in isolation, producing both a topological and temporal estimate of WGD events. Molecular clock approaches to dating WGD have either been flawed by the underlying algorithm [45] or, when more powerful Bayesian uncorrelated methods have been used, by the limited sampling of taxa and a paucity of appropriate fossil calibrations [40]. Furthermore, dating individual gene families does not make best use of information available because individual gene families have low statistical power, yielding imprecise, if not inaccurate, estimates of gene and (by inference) genome duplication dates.

Paralog sets derived from a WGD share the same age and can be combined in a concatenated alignment that is capable of producing far more precise results than any single gene family [46,47]. Precision is not the sole concern, and improved accuracy is achieved by using conservative paleontological constraints on speciation events [48] alongside clock methods that can model both the uncertainty in the fossil evidence and the variation in rates of evolution between genes [46,49]. Box 3 shows a schematic analysis of the WGD present in the ancestor

blue, triplications in red, known autopolyploidy events in yellow, and allopolyploidy events in green. The white bar associated with Caryophyllales represents 26 independent WGD events, some of which are autopolyploidy and some allopolyploidy. Named duplication events are shown alongside their Greek letter. Abbreviations: Camb., Cambrian; Carb., Carboniferous; Ord., Ordovician; Neo., Neogene; Pal., Paleozoic; Sil., Silurian.

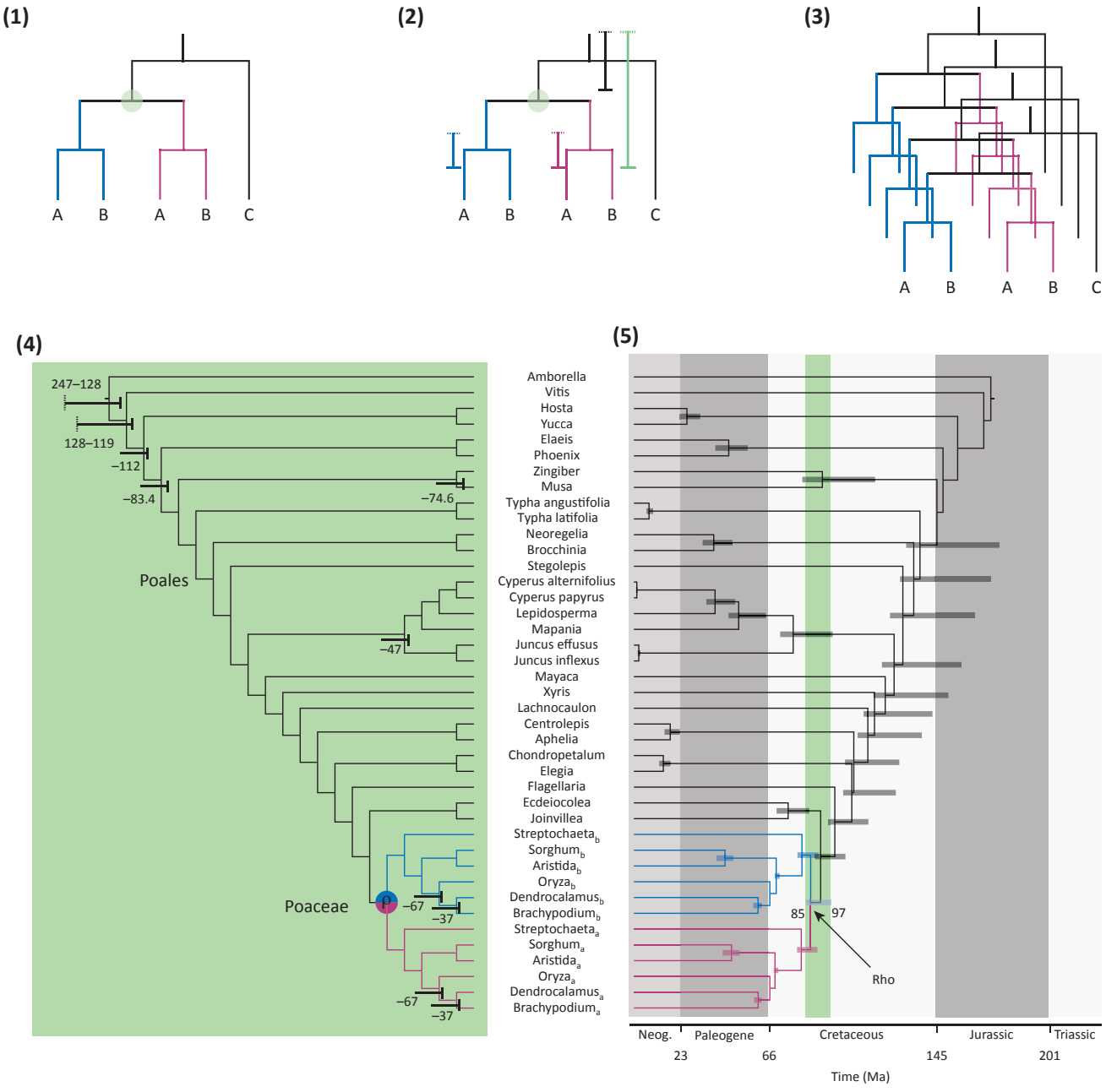


Figure 2. Dating Whole-Genome Duplication (WGD) by Combining Genomics and the Fossil Record. (1) the history of WGD is present in individual gene families. Taxa A and B have undergone a shared duplication event, which taxon C has not. (2) The timing of the duplication is bracketed by the timing of the divergence of A and B and the divergence of A + B and C. These divergence times can be calibrated using distributions between minimum and soft maximum ages. (3) Multiple gene families with a shared WGD signal can be concatenated to maximize the precision of the analysis. (4) Accuracy is achieved through careful appraisal of the fossil record and by modeling uncertainty through soft maximum ages [46,94]. (5) A Bayesian molecular clock analysis reveals that the grass duplication (Rho) occurred at 85–97 Ma (95% highest posterior density, HPD).

Box 3. Dating WGD in Grasses

Syntenic and phylogenomic evidence points towards a WGD event in the ancestor of all extant grasses (Poaceae) [23,30]. The Rho event has previously been dated through phylogenetic bracketing to ~70 Ma [90] and is one of the numerous plant WGDs hypothesized to approximate the K–Pg boundary [91]. We sampled the gene families that were previously shown to retain the signal of the Rho duplication (Figure 2.1) and concatenated them into an alignment (Figure 2.3; see File S2 in the supplemental information online). Fossil evidence constrains the minimum age on speciation nodes (see File S1 in the supplemental information online), and in some cases can be used to apply 'soft' maxima [92] (Figure 2.2). The Late Cretaceous fossil phytolith taxon *Changii indicum* is assigned to the crown group (i.e., the living clade) of the Oryzeae tribe, and provides a minimum age of 66 Ma based on radiometric dating [93–95]. This fossil placement of this fossil is contentious and can be instead used to calibrate the BOP (Bambusoideae, Oryzoideae, Pooideae) + PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) clade of grasses [95]. We applied further fossil constraints and, combined with the concatenated alignment, these calibrations inform a Bayesian molecular clock analysis performed on the fixed topology of McKain *et al*. in the phylogenetic program MCMCTREE [96]. The results predict that the WGD took place in the 97–85 Ma period, and in this case is not compatible with the hypothesis that this event coincides with the K–Pg boundary (Figure 2.5). Abbreviation: ANA grade.

of all grasses (Rho). This event is evident in the genomes and phylomes of multiple extant grass species which, owing to their economic value as food crops, have been well-sampled by sequencing projects [30].

As well as being able to inform on the coincidence of WGD with geological or biogeographic events, these approaches coestimate the timing of duplication alongside the timing of speciation. This allows us to see how early or late WGD occurred relative to the crown (extant) clade and to directly estimate lag between the WGD event and any hypothesized macroevolutionary consequences [46].

Whole Genomes and Diversification

Diversification is one of the most widely proposed consequences of WGD in plants. This relationship has been explored at multiple levels across angiosperms, but support for a correlation remains equivocal [2,29,50,51]. There is little evidence supporting a direct shift in diversification immediately following WGD. Instead, there is some support for the proposed 'WGD lag-time' model, wherein diversification follows WGD – but only after a protracted period of geological time [2]. The lag period has been measured either as a period of absolute time or as an arbitrary measure of time such as the number of nodes separating a WGD event and a subsequent shift in the rate of diversification. When the age of the duplication event and the subsequent speciation events are coestimated, the absolute age and duration of the lag can be estimated directly [46]. Estimates for the timing of the angiosperm-specific genome duplication event imply that it occurred 65–35 Myr before the divergence of crown angiosperms (the living clade of flowering plants), closer to 70 Myr before the radiation of the Mesangiospermae and over 100 Myr before a detectable angiosperm radiation in the fossil record [46,52]. Such an extensive lag raises two questions: first, is it plausible to associate two events that are separated by such a long interval of time? Second, why did the early diverging lineages of angiosperms (the ANA grade: Amborellales, Nymphaeales, and Austrobaileyales) not undergo a similar radiation?

Schranz *et al*. [53] proposed a model in which WGD provides latent evolvability that may be later triggered by a shift in environment and promote diversification. This has been further refined, and several new models have emerged to explain the lag phase, some of which are readily testable. Among these is the suggestion that it is not WGD, but the ensuing process of genome fractionation (or **diploidization**), that may be responsible for diversification. During this process the organism undergoes large-scale genome rearrangements and redundant gene copies are silenced and excised, leading to potentially novel patterns of expression [54]. Most angiosperm lineages have undergone multiple rounds of WGD and exhibit the fastest rate of genome size

evolution among land plants [55], and it has been proposed that their ability to rapidly downsize their genome in the wake of WGD has led to their global dominance [56]. Ferns show a higher rate of genome duplication than angiosperms but appear not to undergo such extensive genome downsizing and are considerably less diverse than angiosperms [33,57]. The observed lag between WGD and diversification in angiosperms may be explained by the period of genome fractionation, although the long-term rate of fractionation is uncertain. It seems appropriate to ask whether the extent or rate of genome reorganization post-WGD correlates with observed shifts in the rate of diversification. The WGD event associated with one of the most dramatic shifts in diversification, the gamma event at the base of the eudicots, involved extensive genome reorganization [25,58]. Speciation post-WGD would lead to fractionation occurring independently in separate lineages, which could explain the differences between lineages that emerge from WGD [54].

In the specific case of autopolyploidy (duplication involving a single parental lineage) the newly duplicated paralogs can pair randomly at meiosis. This pattern of tetrasomic inheritance facilitates ongoing exchange between paralogous chromosomes and may prevent them from diverging until a state of disomic inheritance is restored [59,60]. The period required to attain a state of disomic inheritance could also explain the macroevolutionary lag between WGD and phenotypic evolution. As with the duplication/fractionation model, speciation occurring before the restoration of disomic inheritance will result in independent diploidization of lineages. Robertson *et al*. [59] demonstrated this 'lineage-specific **ohnolog** resolution' (LORe) model in the descendants of the salmonid fish-specific WGD event, and showed that independent diploidization was present in 27% of salmonid paralogs. Although untested in plants, its predictions of a long lag period and disparate evolutionary trajectories suggest that it may also fit the patterns observed after the angiosperm-specific WGD.

The case for a general theory of WGD as an intrinsic driver of diversification is undermined by the multiple cases where WGD does not accompany any shift in diversity. Non-seed-plant lineages, such as paleopolyploid mosses and horsetails, remain species-poor despite repeated duplications [7,10]. This can be partly reconciled by the differing rates of genome downsizing and rearrangement exhibited by these clades relative to angiosperms. However, further research on the mechanisms for rapidly altering genome structure is required. Beyond plants and, in particular, among teleost fish, the paleontological record shows no evidence in support of a role for WGD in directly promoting diversification [61]. There is some evidence supporting a direct role for WGD in promoting diversity in yeasts where reciprocal gene loss can lead to reproductive isolation [62], although on a macroevolutionary scale this effect is small [63].

WGD and Morphological Innovation

The link between WGD and morphological evolution in plants has remained both pervasive and speculative [1,64]. Some have proposed that polyploids may survive and evolve in extreme or marginal habitats, allowing them a competitive advantage over their parent species at range margins [65]. However, the range of many extant polyploids does not exceed that of their parents [66], while genes related to stress tolerance appear to have evolved via tandem duplication rather than by WGD [67,68]. The evolution of morphological diversity, like species diversity, may also require a lag phase. For selection to act on innovation, developmental robustness is required [69], and hence it is possible that morphological diversification may occur only after a period of developmental lability. At the genetic level, WGD may free a lineage from the constraints of purifying selection and allow genes to take on new functions [1], such as through **neofunctionalization** and **subfunctionalization**. At the phenotypic level this may allow the evolution of novel forms and body plans. Indeed, formative innovations within the plant kingdom have been

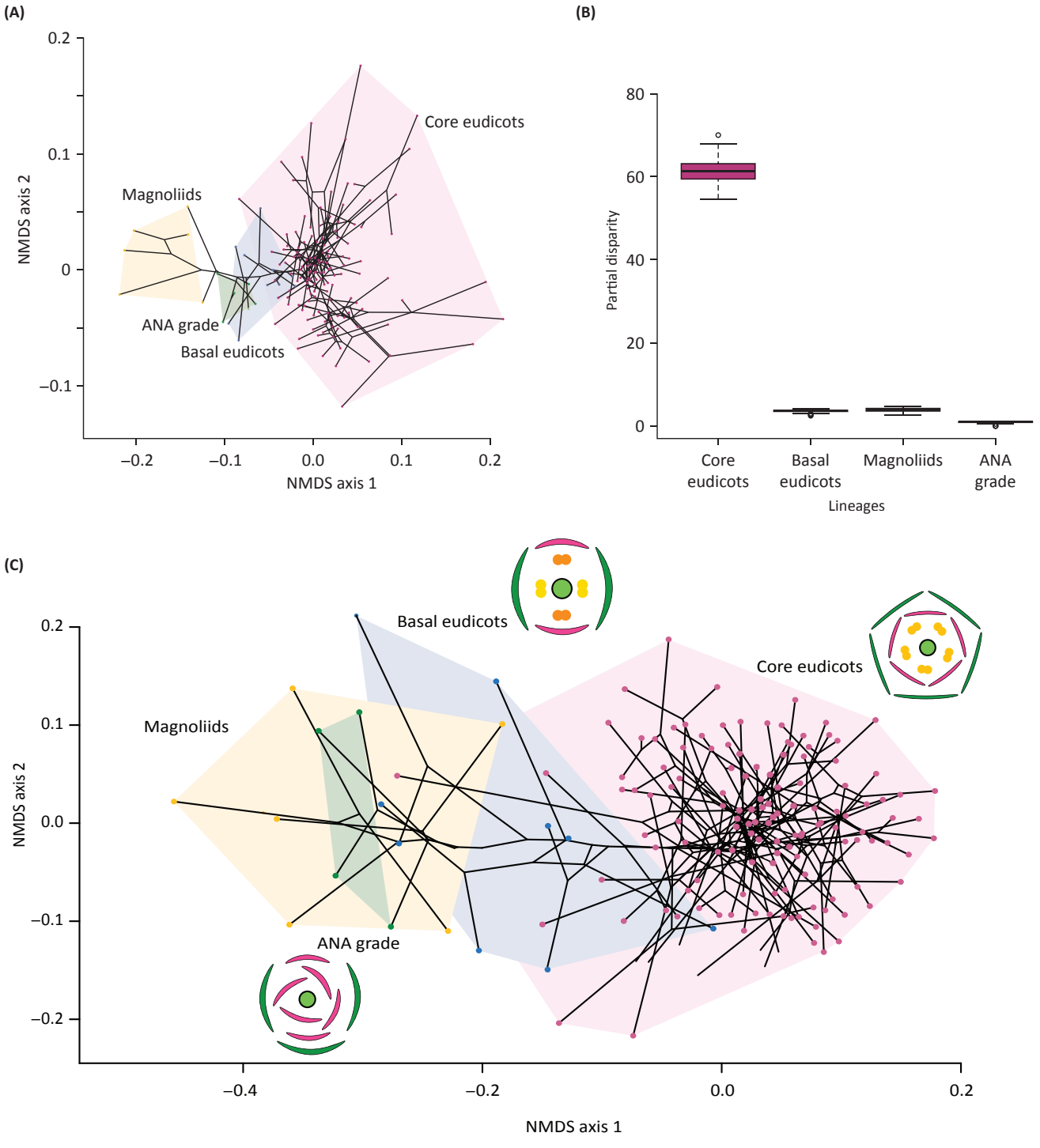


Figure 3. Phenotypic Evolution in the Wake of the Gamma Triplication Which Occurred before the Evolution of the Core Eudicots. (A) An empirical morphospace based on a morphological matrix [85]. Morphological characters form the basis of a distance metric (Gower's index) which was subjected to non-metric multidimensional scaling (NMDS) to display variation in two axes. A consensus phylogeny is mapped onto the morphospace (see File S4 in the supplementary

(See figure legend on the bottom of the next page.)

associated with the expansion of families of regulatory genes [70,71]. Patterns of gene retention post-WGD are not random, and in repeated cases genes encoding proteins that function as part of networks and signaling cascades are retained preferentially [72–74]. This has been explained in terms of dosage balance and the need to maintain stoichiometric ratios of proteins within the cell [75,76]. The dosage-balance hypothesis is exemplified during the diploidization process in allopolyploids, where exchanges can occur between homoeologous chromosomes of subgenomes [77]. These exchanges can result in novel gene expression and gene copy number [78], but can also result in the deleterious loss of chromosome regions or entire chromosomes. Homoeologous compensation has been proposed as a mechanism to prevent dosage imbalances, and has been demonstrated to lead to increased phenotypic variation in newly synthesized allopolyploids [77]. The dosage-balance hypothesis does not predict the evolution of morphological diversity until such constraints are relaxed and retained paralogs are selected to evolve new functions [14,79]. These constraints may relax under different selection pressures, although a quantitative model of compensatory drift has also been proposed [80]. Compensatory drift is the process whereby paralogs are initially retained due to dosage sensitivity, but over time the expression levels of the individual genes drift until one paralog is free of the dosage-dependent constraint [80]. This model not only provides a mechanism for neofunctionalization to arise from a state of dosage balance but also a potential explanation for the emergence of evolutionary novelty after prolonged periods of evolutionary time.

It is difficult to ascribe adaptive evolution to WGD, especially with ancient events. The link between WGD and novelty has been elegantly shown in the glucosinolate pathway in Brassicales [4]. This gene family has expanded over several rounds of WGD and is involved in plant–herbivore interactions. It has also been proposed that gene families underpinning floral patterning expanded during the angiosperm-specific WGD [71]. These genes are implicated in the origin and diversification of the flower, a structure that has shaped recent plant and animal evolution [81]. The evolution of pentamerous flowers in the core eudicots also coincides with a genome triplication (gamma, Figure 1) [25,82]. The coincidence of the gamma event with this major synapomorphy, a large increase in the rate of diversification, and extensive genome reorganization [58] makes it a tantalizing system in which to investigate the link between WGD and morphological evolution.

Regulatory gene retention and large shifts in their transcription patterns suggest a role for WGD in the evolution of eudicot floral diversity [82]. To make such a hypothesis testable, the increase in phenotypic complexity must be quantified for comparative analysis [83]. To achieve this we can borrow from paleontology, which has a strong tradition in comparative analysis of phenotype through multivariate statistics – manifest as **morphospace** analyses. The hypothesis that WGD drives innovation would predict that events coincide either with movement to a new 'island' within phenotype 'design space' or with continued expansion of an existing island. These predictions can be tested explicitly with datasets that use discrete phenotypic characters to describe the traits that unify and distinguish taxa [84]. For example, one can characterize the disparity of extant angiosperms to test the hypothesis that the gamma triplication event coincides with an increase in phenotypic diversity. To do this we used a morphological dataset that captures the disparity of early angiosperms, basal eudicots, and core eudicots [85] (see File S3 in the supplemental information online). We used these data to calculate the dissimilarity between each taxa, as measured using Gower's dissimilarity metric [86]. To visualize this dissimilarity we performed non-

information online). (B) The contribution to total disparity (partial disparity) of each clade calculated from distance matrix (1000 bootstrap replicates) [105]. (C) A morphospace constructed from floral characters. Major trends in floral evolution are displayed next to the lineages; spiral phyllotaxis is present in early angiosperms, dimerous flowers are common among basal eudicots, and the pentamerous flower is associated with the core eudicots. Abbreviation: ANA grade, from Amborellales, Nymphaeales, and Austrobaileyales.

Box 4. Duplication and Disparity in the Conifers

Some explosive WGD events, such as that associated with the core eudicots, coincide with rapid diversification and an increase in phenotypic variation. However, many WGD events in species-poor lineages are not closely associated with macroevolutionary phenomena. Most conifers are thought to have undergone at least two rounds of WGD during their evolution, one shared with seed plants, and then two lineage-specific events on the branches leading to Pinaceae and Cupressophytes [22]. Preliminary analyses of diversity and disparity in the pines indicate a rapid increase in phenotypic variance during the Late Jurassic and Early Cretaceous [83], and the Pinaceae occupy a highly distinct area of morphospace (Figure I). This provides some corroborative support for the hypothesis that WGD resulted in phenotypic variation among conifers during their early evolution. However, the age of the pine WGD is currently estimated at between 342 and 200 Ma [22] (Figure I); given such uncertainty it is not presently possible to link WGD to the shift in phenotypic disparity. This example highlights the need to employ methods that can accurately and precisely estimate the timing of WGD events because a temporal framework is essential for testing macroevolutionary hypotheses [46].

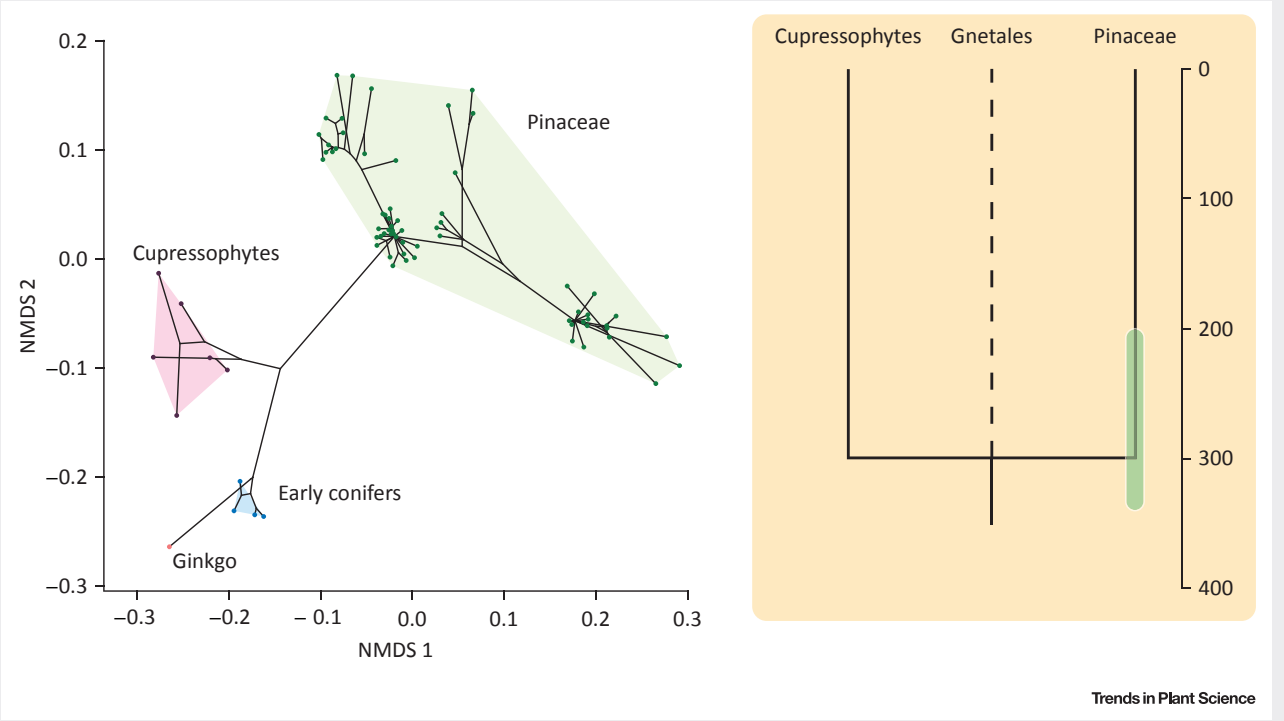


Figure I. Evolution in Pinaceae. An empirical morphospace of Pinaceae and relatives built from phenotypic characters [106] (see File S5 in the supplemental information online) which formed the basis of a distance matrix (Gower's index) that was subjected to non-metric multidimensional scaling (NMDS). A consensus phylogeny is mapped onto the morphospace (see File S6 in the supplemental information online). The uncertainty of both the relative (phylogenetic) and absolute timing of the event limits our understanding of the consequences because the position of the Gnetales remains contentious and the current estimate for the age of the WGD spans over 100 Myr.

metric multidimensional scaling, a non-metric ordination method that summarizes variation over a specified number of axes – in this instance, two. The result is presented in Figure 3, which shows that the core eudicots occupy a far greater area of morphospace than the basal eudicots. Furthermore, relative to other early diverging lineages of angiosperms, they occupy the largest proportion of morphospace (partial disparity, Figure 3B). In addition, we subsampled the character matrix for floral characters only, relating specifically to the gamma-derived hypothesis (Figure 3C). The resulting morphospace shows less separation between the lineages, but core eudicots still occupy the largest area and, therefore, exhibit the greatest variation. The construction of a morphospace can be subjective in that it is dependent on the choice of taxa and characters – but there is strong evidence to suggest that the gamma triplication coincides with the rapid evolution of morphological disparity among eudicots. A comparable analysis of the impact of WGD in pines finds support for increased variance in morphospace occupation, but gross uncertainty in the estimate of the timing of WGD relative to the age of the disparate clade undermines the hypothesis of a causal link (Box 4).

Quantifying phenotypic evolution across multiple lineages will be instrumental in understanding the role of WGD in the evolution of morphologic complexity. The inclusion of fossil taxa and recent methods used to estimate disparity through time may allow us to measure the tempo of phenotypic evolution post-WGD. The impact of key innovations that are attributed to WGD can be tested by considering their impact on the shape of a morphospace or whether the innovation has resulted in diversification. A further question arises as to what degree WGD is essential for the appearance of major innovations. For example, the origin of seed and flowering plants coincides with a WGD event, but, arguably, a greater number of characters unite the vascular plants whose origin was independent of any known WGD event [87]. While it is plausible that saltational evolution has been caused by WGD in the plant kingdom [88], phenotypic complexity may also arise through the evolution more nuanced *trans*- and *cis*-acting regulation [89].

Concluding Remarks

WGD is associated with a macroevolutionary outcome in some but not all lineages, and it remains unclear how and why this is the case. As the number of identified WGD events in plant evolutionary history increases, there is an ever-growing need for a general theory on the role of WGD in macroevolution. However, to establish whether WGD is a class of event with characteristic and predictable outcomes, further work will be necessary to place, both relatively and absolutely, each event in time. There are many outstanding questions to be answered, but a precise temporal framework forms the basis for tests that can quantify any macroevolutionary consequences and inform and refine hypotheses about the relationship between WGD, diversification, and morphological evolution. Plants are the best system in which to elucidate the effects of WGD because of the prevalence of these genomic events in plant phylogeny. This will be crucial as we seek to explain the consequences beyond any single event and, given the role that genome duplication has had in the evolution of many crop species, being able to make general predictions about the outcome of WGD is of crucial interest (see Outstanding Questions).

Outstanding Questions

Questions remain about the absolute timing of many of the identified WGD events among plants – of particular interest in the clustering of events around the K–Pg boundary.

The origin of duplication events is important – it has implications for both the timing and evolutionary consequences.

Is morphological evolution accelerated in the wake of WGD, and what impact has WGD had on the plant morphospace?

Disparate outcomes between lineages, in terms of morphology and diversity, still require investigation.

Acknowledgments

We thank Simon Hiscock, Kyle Martin, Mark Puttick, Harald Schneider, and Ziheng Yang for discussion. J.E.C. is funded by a PhD studentship from the Biotechnology and Biological Sciences Research Council (BBSRC; swDTP); P.C.J.D. is funded by grants from the Natural Environment Research Council (NERC; NE/N002067/1) and the BBSRC (BB/N000919/1).

Appendix A Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tplants.2018.07.006.

References

1. Ohno, S. (1970) *Evolution by Gene Duplication*, Springer Science Business Media
2. Tank, D.C. *et al.* (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467
3. Ren, R. *et al.* (2018) Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11, 414–428
4. Edger, P.P. *et al.* (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366
5. Barker, M.S. *et al.* (2016) Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am. J. Bot.* 103, 1203–1211
6. Smith, S.A. *et al.* (2018) Disparity, diversity, and duplications in the Caryophyllales. *New Phytol.* 217, 836–854
7. Vanneste, K. *et al.* (2015) Horsetails are ancient polyploids: evidence from *Equisetum giganteum*. *Plant Cell* 27, 1567–1578
8. Lohaus, R. and Van de Peer, Y. (2016) Of dups and dinos: evolution at the K/Pg boundary. *Curr. Opin. Plant Biol.* 30, 62–69
9. Fawcett, J.A. *et al.* (2009) Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl. Acad. Sci.* 106, 5737–5742
10. Devos, N. *et al.* (2016) Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytol.* 211, 300–318
11. Cui, L. *et al.* (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749
12. Walker, J.F. *et al.* (2017) Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *Am. J. Bot.* 104, 858–867

13. Xiang, Y. *et al.* (2017) Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* 34, 262–281
14. Conant, G.C. *et al.* (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* 19, 91–98
15. Donoghue, P.C. and Purnell, M.A. (2005) Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.* 20, 312–319
16. Schwager, E.E. *et al.* (2017) The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62
17. Marcet-Houben, M. and Gabaldón, T. (2015) Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* 13, e1002220
18. Li, Z. *et al.* (2018) Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci.* 115, 4713–7718
19. Arrigo, N. and Barker, M.S. (2012) Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* 15, 140–146
20. Soltis, D.E. *et al.* (2011) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al*. (2011). *New Phytol.* 202, 1105–1117
21. Jiao, Y. *et al.* (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 91–100
22. Li, Z. *et al.* (2015) Early genome duplications in conifers and other seed plants. *Sci. Adv.* 1, e1501084
23. Jiao, Y. *et al.* (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* 26, 2792–2802
24. Jaillon, O. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467
25. Jiao, Y. *et al.* (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* 13, R3
26. Huang, C.-H. *et al.* (2016) Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol. Biol. Evol.* 33, 2820–2835
27. Kagale, S. *et al.* (2014) Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* 26, 2777–2791
28. Cannon, S.B. *et al.* (2015) Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32, 193–210
29. Estep, M.C. *et al.* (2014) Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci.* 111, 15149–15154
30. McKain, M.R. *et al.* (2016) A phylogenomic assessment of ancient polyploidy and genome evolution across the poales. *Genome Biol. Evol.* 8, 1150–1164
31. Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
32. Vanneste, K. *et al.* (2013) Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* 30, 177–190
33. Clark, J. *et al.* (2016) Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol.* 210, 1072–1082
34. Tiley, G.P. *et al.* (2016) Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. *Genome Biol. Evol.* 8, 1023–1037
35. Tang, H. *et al.* (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18, 1944–1954
36. Lyons, E. *et al.* (2008) Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with Rosids. *Plant Physiol.* 148, 1772–1781
37. Edger Patrick, P. *et al.* (2018) Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am. J. Bot.* 105, 463–469
38. Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
39. Blanc, G. and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678
40. Vanneste, K. *et al.* (2014) Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* 24, 1334–1347
41. Guan, R. *et al.* (2016) Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* 5, 49
42. Roodt, D. *et al.* (2017) Evidence for an ancient whole genome duplication in the cycad lineage. *PLoS One* 12, e0184454
43. Morris, J.L. *et al.* (2018) The timescale of early land plant evolution. *Proc. Natl. Acad. Sci.* 115, E2274–E2283
44. Huerta-Cepas, J. *et al.* (2014) PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* 42, D897–D902
45. Ruprecht, C. *et al.* (2018) Revisiting ancestral polyploidy in plants. *Sci. Adv.* 3, e1603195
46. Clark, J.W. and Donoghue, P.C.J. (2017) Constraining the timing of whole genome duplication in plant evolutionary history. *Proc. Biol. Sci.* 284, 20170912
47. Macqueen, D.J. and Johnston, I.A. (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. Biol. Sci.* 281, 20132881
48. Parham, J.F. *et al.* (2012) Best practices for justifying fossil calibrations. *Syst. Biol.* 61, 346–359
49. Warnock, R.C. *et al.* (2015) Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. Biol. Sci.* 282, 20141013
50. Kellogg, E.A. (2016) Has the connection between polyploidy and diversification actually been tested? *Curr. Opin. Plant Biol.* 30, 25–32
51. Soltis, D.E. *et al.* (2009) Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348
52. Silvestro, D. *et al.* (2015) Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytol.* 207, 425–436
53. Schranz, M.E. *et al.* (2012) Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr. Opin. Plant Biol.* 15, 147–153
54. Dodsworth, S. *et al.* (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* 180, 1–5
55. Puttick, M.N. *et al.* (2015) Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proc. Biol. Sci.* 282, 20152289
56. Simonin, K.A. and Roddy, A.B. (2018) Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biol.* 16, e2003706
57. Wood, T.E. *et al.* (2009) The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci.* 106, 13875–13879
58. Wang, Y. *et al.* (2016) Large-scale gene relocations following an ancient genome triplication associated with the diversification of core eudicots. *PLoS One* 11, e0155637
59. Robertson, F.M. *et al.* (2017) Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* 18, 111
60. Martin, K.J. and Holland, P.W.H. (2014) Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication. *Mol. Biol. Evol.* 31, 2592–2611

61. Laurent, S. *et al.* (2017) No evidence for the radiation time lag model after whole genome duplications in Teleostei. *PLoS One* 12, e0176384
62. Maclean, C.J. and Greig, D. (2011) Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. *Evolution* 65, 932–945
63. Muir, C.D. and Hahn, M.W. (2015) The limited contribution of reciprocal gene loss to increased speciation rates following whole-genome duplication. *Am. Nat.* 185, 70–86
64. Crow, K.D. and Wagner, G.P. (2006) What is the role of genome duplication in the evolution of complexity and diversity? *Mol. Biol. Evol.* 23, 887–892
65. Stebbins, G.L. (1947) Types of polyploids: their classification and significance. *Adv. Genet.* 1, 403–429
66. Glennon, K. *et al.* (2014) Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecol. Lett.* 17, 574–582
67. Panchy, N. *et al.* (2016) Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316
68. Hanada, K. *et al.* (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003
69. Melzer, R. and Theißen, G. (2016) The significance of developmental robustness for species diversity. *Ann. Bot.* 117, 725–732
70. Rensing, S.A. (2014) Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* 17, 43–48
71. Chanderbali, A.S. *et al.* (2016) Evolving ideas on the origin and evolution of flowers: new perspectives in the genomic era. *Genetics* 202, 1255–1265
72. Seoighe, C. and Gehring, C. (2004) Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* 20, 461–464
73. Qiao, X. *et al.* (2018) Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). *Front. Plant Sci.* 9, 161
74. Veron, A.S. *et al.* (2007) Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. *Mol. Biol. Evol.* 24, 670–678
75. Veitia, R.A. *et al.* (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 24, 390–397
76. Birchler, J.A. and Veitia, R.A. (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14746–14753
77. Xiong, Z. *et al.* (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7908–7913
78. Lloyd, A. *et al.* (2017) Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytol.* 217, 367–377
79. Freeling, M. and Thomas, B.S. (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* 16, 805–814
80. Thompson, A. *et al.* (2016) Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. *Genetics* 202, 765–774
81. Fernández-Mazuecos, M. and Glover, B.J. (2017) The evo-devo of plant speciation. *Nat. Ecol. Evol.* 1, 110
82. Chanderbali, A.S. *et al.* (2017) Evolution of floral diversity: genomics, genes and gamma. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20150509
83. Oyston, J.W. *et al.* (2016) Why should we investigate the morphological disparity of plant clades? *Ann. Bot.* 117, 859–879
84. Hetherington, A.J. *et al.* (2015) Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology* 58, 393–399
85. Nandi, O.I. *et al.* (1998) A combined cladistic analysis of angiosperms using rbcL and non-molecular data sets. *Ann. Mo. Bot. Gard.* 85, 137–212
86. Gower, J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27, 857–871
87. Banks, J.A. *et al.* (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963
88. Minelli, A. (2018) *Plant Evolutionary Developmental Biology: The Evolvability of the Phenotype,* Cambridge University Press
89. Chen, C.-Y. *et al.* (2012) Lengthening of 3′UTR increases with morphological complexity in animal evolution. *Bioinformatics* 28, 3178–3181
90. Paterson, A. *et al.* (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9903–9908
91. Van de Peer, Y. *et al.* (2017) The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424
92. Donoghue, P.C. and Benton, M.J. (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* 22, 424–431
93. Prasad, V. *et al.* (2011) Late cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nat. Commun.* 2, 480
94. Iles, W.J.D. *et al.* (2015) Monocot fossils suitable for molecular dating analyses. *Bot. J. Linnean Soc.* 178, 346–374
95. Christin, P.-A. *et al.* (2014) Molecular dating, evolutionary rates, and the age of the grasses. *Syst. Biol.* 63, 153–165
96. Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591
97. Steige, K.A. and Slotte, T. (2016) Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. *Curr. Opin. Plant Biol.* 30, 88–93
98. Spoelhof, J.P. *et al.* (2017) Pure polyploidy: closing the gaps in autopolyploid research. *J. Syst. Evol* 55, 340–352
99. Bottani, S. *et al.* (2018) Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* 23, 393–402
100. Garsmeur, O. *et al.* (2014) Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* 31, 448–454
101. Julca, I. *et al.* (2018) Phylogenomics of the olive tree (*Olea europaea*) disentangles ancient allo- and autopolyploidizations in Lamiales. *BMC Biol.* 16, 15
102. Thomas, G.W.C. *et al.* (2017) Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Syst. Biol.* 66, 1007–1018
103. Doyle, J.J. and Egan, A.N. (2010) Dating the origins of polyploidy events. *New Phytol.* 186, 73–85
104. Rothfels, C.J. *et al.* (2015) Natural hybridization between genera that diverged from each other approximately 60 million years ago. *Am. Nat.* 185, 433–442
105. Guillerme, T. (2015) dispRity: a modular R package for measuring disparity. *Methods Ecol. Evol.* 9, 1755–1763
106. Smith, S.Y. *et al.* (2017) A new species of *Pityostrobus* (Pinaceae) from the Cretaceous of California: moving towards understanding the Cretaceous radiation of Pinaceae. *J. Syst. Palaeontol.* 15, 69–81

https://www.nature.com/articles/s41559-018-0644-x

Evolution of metazoan morphological disparity

Bradley Deline[a,1], Jennifer M. Greenwood[b], James W. Clark[b], Mark N. Puttick[b,c], Kevin J. Peterson[d], and Philip C. J. Donoghue[b,1]

[a]Department of Geoscience, University of West Georgia, Carrollton, GA 30118; [b]School of Earth Sciences, University of Bristol, BS8 1TQ Bristol, United Kingdom; [c]Department of Biology and Biochemistry, University of Bath, BA2 7AY Bath, United Kingdom; and [d]Department of Biological Science, Dartmouth College, Hanover, NH 03755

Edited by Neil H. Shubin, The University of Chicago, Chicago, IL, and approved August 7, 2018 (received for review June 26, 2018)

The animal kingdom exhibits a great diversity of organismal form (i.e., disparity). Whether the extremes of disparity were achieved early in animal evolutionary history or clades continually explore the limits of possible morphospace is subject to continuing debate. Here we show, through analysis of the disparity of the animal kingdom, that, even though many clades exhibit maximal initial disparity, arthropods, chordates, annelids, echinoderms, and mollusks have continued to explore and expand the limits of morphospace throughout the Phanerozoic, expanding dramatically the envelope of disparity occupied in the Cambrian. The "clumpiness" of morphospace occupation by living clades is a consequence of the extinction of phylogenetic intermediates, indicating that the original distribution of morphologies was more homogeneous. The morphological distances between phyla mirror differences in complexity, body size, and species-level diversity across the animal kingdom. Causal hypotheses of morphologic expansion include time since origination, increases in genome size, protein repertoire, gene family expansion, and gene regulation. We find a strong correlation between increasing morphological disparity, genome size, and microRNA repertoire, but no correlation to protein domain diversity. Our results are compatible with the view that the evolution of gene regulation has been influential in shaping metazoan disparity whereas the invasion of terrestrial ecospace appears to represent an additional gestalt, underpinning the post-Cambrian expansion of metazoan disparity.

Metazoa | disparity | evolution | morphology | Cambrian explosion

The diversity of animal organismal form (i.e., disparity) is decidedly nonrandom; members of one phylum share "bodyplan" characteristics distinct from those of other phyla, suggesting that only a very small subset of the universe of possible bodyplans has been realized. Paleontological analyses have suggested that the limits on organismal disparity were realized early in animal evolutionary history (1–3), inspiring the view that fundamental innovation has been precluded subsequently by gene regulatory developmental constraints (4–6), and that the evolutionary processes underlying the emergence of animals are nonuniformitarian (7, 8). However, this perspective is based largely on the timing of appearance of Linnean rank taxa in the fossil record (8, 9), assuming they provide an effective proxy for organismal disparity. Attempts to capture disparity by using morphometry have borne out the hypothesis of maximal initial disparity (3, 10), but this approach is limited practically to analysis at low taxonomic levels and it is not clear that the results can be generalized to higher taxa, including the phylum and kingdom levels. Here we attempt to map metazoan disparity within an empirical morphospace based on a large sampling of discrete characters from across the breadth of extant metazoan diversity. We use this to explore the impact of extinction on morphospace occupation and the relationship between organismal disparity and other phenomena such as complexity, body size, diversity, and Linnean rank. We then undertake quantitative tests of hypotheses of causality, including random variation, genome size, protein diversity, and gene regulatory complexity.

Mapping Metazoan Morphospace

The construction of a morphospace is dependent on methodology and the selection of relevant features. However, with a significantly large data source, the distances among taxa within morphospace will begin to approximate the evolutionary scale of the differences. Spatial landmark analysis is precluded at high taxonomic rank such as phylum because, by definition, phyla share few morphological homologies. Discrete characters provide a suitable alternative given that there are no practical limits to their scalability (2, 3), and comparative analyses have shown that continuous and discrete character datasets can capture the same phenomenon (11–14). The use of discrete characters produces results that have nonmetric properties (15–17), but this approach can and has been used to elucidate broad patterns of similarities and clustering within multidimensional space (18), particularly in formulating the hyoptheses we seek to test. To test between competing hypotheses for the evolution of disparity—whether the limits of disparity were established early or have continued to expand throughout evolutionary history—we compiled a cladistic character matrix derived from Peter Ax's *Phylogenetic System* (19–21). This constitutes a single, densely sampled synthetic overview of character distribution among metazoans, including all phyla, by an individual who was not a taxonomic specialist in any of the groups that could perhaps be considered overrepresented in the dataset. Ax's taxonomic sampling is not uniform across metazoans, but the number of characters and taxa within a phylum is representative of its intraphylum diversity (ref. 22; Spearman's correlation, $\rho = 0.821$, $P < 0.001$). Therefore, we do not consider that any clades are

Significance

We attempt to quantify animal "bodyplans" and their variation within Metazoa. Our results challenge the view that maximum variation was achieved early in animal evolutionary history by nonuniformitarian mechanisms. Rather, they are compatible with the view that the capacity for fundamental innovation is not limited to the early evolutionary history of clades. We perform quantitative tests of the principal hypotheses of the molecular mechanisms underpinning the establishment of animal bodyplans and corroborate the hypothesis that animal evolution has been permitted or driven by gene regulatory evolution.

Author contributions: B.D., K.J.P., and P.C.J.D. designed research; B.D., J.M.G., J.W.C., M.N.P., K.J.P., and P.C.J.D. performed research; M.N.P. contributed new reagents/analytic tools; B.D., J.M.G., and J.W.C. analyzed data; and B.D., J.M.G., J.W.C., M.N.P., K.J.P., and P.C.J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: bdeline@westga.edu or phil.donoghue@bristol.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810575115/-/DCSupplemental.

significantly underrepresented or overrepresented in the dataset. We coded 1,767 characters for 212 extant, terminal taxa, including 34 animal phyla, most commonly down to the Linnean rank of order (*SI Appendix*, Table S1). The characters encompass all aspects of morphology (cellular, developmental, sexual, and skeletal and soft-tissue anatomy), including those minimally defining each clade, comprising 915 characters that are shared among the operational taxa (homologies and homoplasies) and 852 unique (i.e., autapomorphic) characters (*SI Appendix*, Table S2 and Dataset S1).

We mapped the relationships between features to identify characters that are nonapplicable rather than absent, which can be differentiated analytically by using Gower's similarity metric (23–25). We subjected these data to a nonmetric multidimensional scaling (NMDS) analysis, a noneigenvector-based multivariate method that attempts to optimize the fit between the data and a preselected number of axes (Fig. 1). The use of a nonmetric ordination technique has several advantages (e.g., the ability to handle large amounts of absent data), but the distances between organisms may not be directly Euclidean, which may alter measurements of disparity. To control for the choice of ordination technique, we repeated all analyses by using principal coordinate analysis. The choice of ordination had no impact on the morphospace or any of the presented results (*SI Appendix*, Fig. S1). The distances between taxa using linear (i.e., principal coordinate analysis) and nonlinear (i.e., NMDS) methods are strongly and linearly correlated (Mantel test, $R^2 = 0.9764$, $P = 0.001$), indicating that, even though the NMDS ordination is built by using a nonlinear method, it has linear properties. Absolute distances within the space can still be subject to nonmetric artifacts such that the distances between taxa should be taken as a qualitative metric of the overall distribution of metazoan morphologic diversity. An analysis of the stress (representing the goodness of fit) indicates that the majority of variance in the data is captured by the first two axes (*SI Appendix*, Fig. S2).

Concerned that different treatments of nonapplicable data (25) could significantly alter the structure of the ordination, we structured our data to reflect different coding strategies (Fig. 2). Treating nonapplicable characters as absent (Fig. 2*A*) or missing (Fig. 2*B*) produced statistically similar ordinations to the main analysis (Fig. 1) in which inapplicable characters are assigned a distinct state conferring distance (Mantel test, $R = 0.907$, $P = 0.001$). However, these differ in the (relative) displacement of the nonmetazoan eukaryotic outgroup, Porifera and Placozoa, into the central area of morphospace (Fig. 2*A*). Treating nonapplicable characters as absent or missing also increased intraphylum disparity at the cost of interphylum disparity (Fig. 2 *A* and *B*). However, the similarity in results from the different treatments of the data indicate the strength of the underlying structure of the data. These results indicate that the structure of disparity is robust to ordination and coding strategies.

The position of taxa based on the first two axes is presented in Fig. 1*A*. Most of this variation is based on shared characters, as analysis of a dataset excluding autapomorphies has no significant impact on the structure of the morphospace ($R^2 > 0.92$, $P = 0.001$; *SI Appendix*, Fig. S3). Phyla differ dramatically in the position and areal extent of their envelope of disparity. Although most taxa are clustered along both axes, the nonmetazoan eukaryote outgroup and Porifera plot separately from eumetazoans, principally for their lack of shared eumetazoan characters, even though neither of these groups occupy a large area of morphospace. Chordata (Fig. 1*B*), Arthropoda (Fig. 1*C*), and, to a lesser extent, Annelida, Echinodermata, and Mollusca, are much more disparate, each occupying larger ranges of morphospace than all other phyla combined, and defining the extremities of morphospace on the two principal axes. A Q-mode analysis of the distribution of characters (Fig. 2*C*) shows that the characters that describe intraphylum features load at the extremities of both axes. Superphylum- and phylum-level characters occupy approximately one fourth of the area of the lower-level characters and vary primarily along the first axis.

From Modern to Historical Disparity

Central to the thesis of maximal initial disparity in animal evolution was the discovery of distinct bodyplans among Cambrian Konservat–Lagerstätten (1), which were assigned historically to extinct phyla, classes, or orders. Thus, by comparing only living taxa, it could be argued that we have captured only net historical disparity. Therefore, we coded a phylogenetically diverse and representative sample of Cambrian taxa, principally the earliest representatives of ordinal level clades (26). This entailed coding 70 fossil taxa for the existing character set and adding 111 mostly autapomorphic characters. Coding these fossil taxa was potentially problematic in that most of the characters (54.1%) are not preserved, and therefore unknown. On average, only 8.6% of the characters were coded as applicable, resulting in the fossil taxa appearing more constrained and skewed toward lower values on the second axis, making the Cambrian taxa appear less complex (Fig. 2*D*); we interpret this result as an artifact of the great volume of data missing for the fossils. There are two possible solutions to accommodating fossil species. One approach is to subsample our dataset for fossilizable characters based on known examples of fossilized features or the anatomical nature of the character (1,000 characters). NMDS analysis of this subsampled dataset results in a plot of morphospace occupation with the same broad structure (Mantel test, $R = 0.974$, $P = 0.001$; Fig. 2*E*) as that recovered from analysis of the entire dataset (Fig. 1*A*). However, the resulting morphospace accentuates the relative disparity of vertebrates and arthropods while diminishing the relative disparity of all other phyla (nonbilaterians especially), individually and in combination, exaggerating the significance of skeletal and gross anatomical characters that are fossilized in instances of routine and exceptional preservation. Few of these characters are representative of bodyplans more generally, which are defined on the basis of soft-tissue, cellular, tissue, organ, and developmental characters that are not usually fossilized. Thus, restricting the analysis to only fossilizable characters cannot be considered to capture organismal disparity within metazoans in any meaningful way. These results are of concern because they suggest that the majority of disparity analyses, which have been applied principally to fossil groups, may have limited inferential power. This is because they are constrained to characterization of fossilizable characters that may be otherwise unrepresentative of phenotypic evolution.

An alternative approach to including fossil species exploits their known phylogenetic position among living and fossil relatives to infer character states that are lost during fossilization. There are obviously assumptions inherent in inferring missing data, including missing secondary reversals in soft tissues, the potential of differential evolutionary rates between preservable and nonpreservable characters, or limiting the coded fossil autapomorphies to preservable characteristics. However, given the rarity of reversals of superphylum-level nonfossilizable characters in extant taxa and the observation that autapomorphies contribute little to the construction of the morphospace (*SI Appendix*, Fig. S3), these assumptions are likely to have a minor impact on the projection of fossil taxa into the morphospace defined by the living species. The approach of inferring missing data likely strengthens the phylogenetic signal in the morphospace. However, a comparison with the taphonomically culled dataset (Fig. 2*E*) indicates a similar and robust placement of the fossil taxa within morphospace.

To implement this approach, we derived a consensus, time-scaled phylogenetic tree for the operational taxa, which differed from Ax's original phylogenetic hypothesis (*SI Appendix*, Figs.

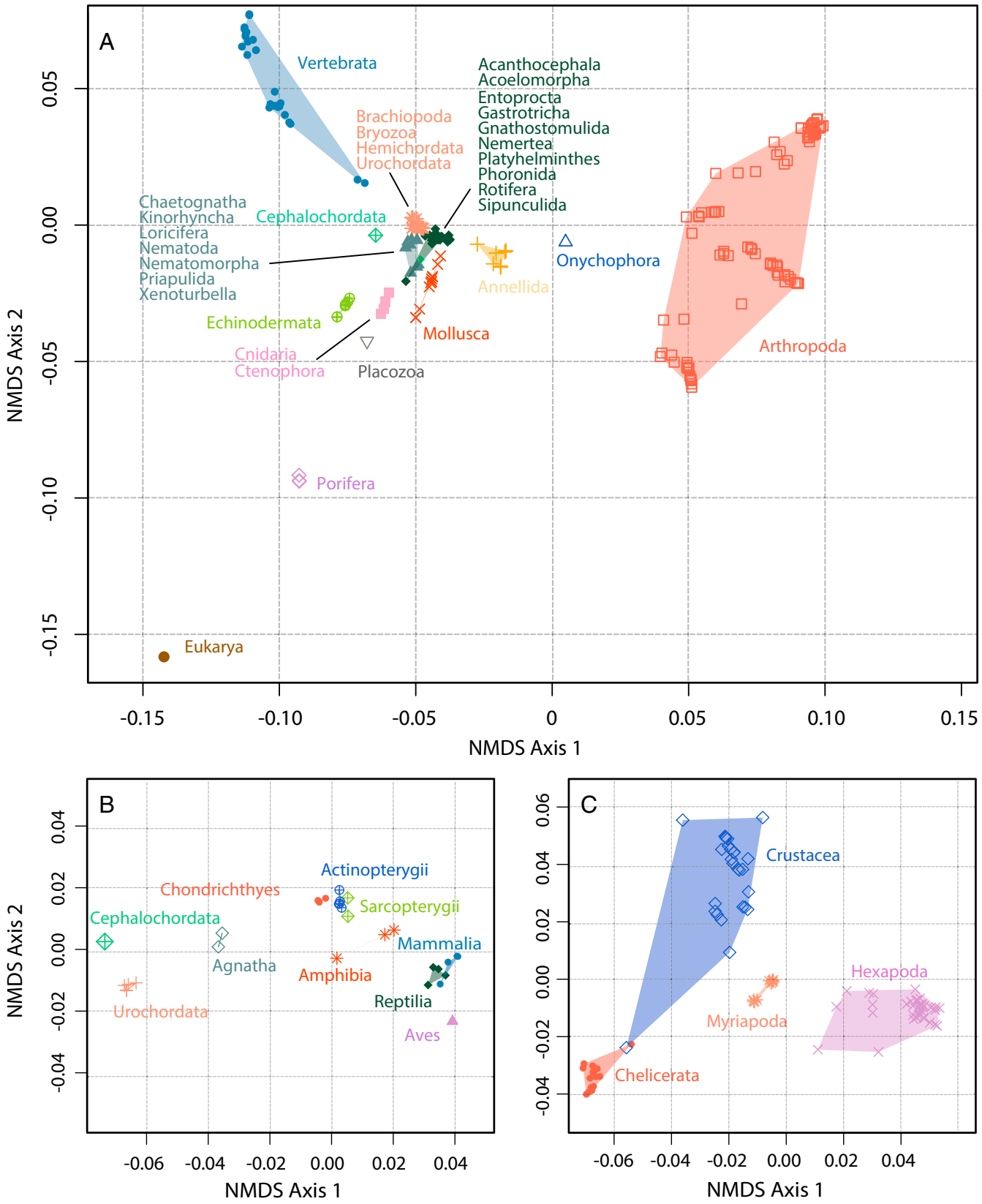


Fig. 1. (*A*) Morphospace encompassing Ax's 212 operational taxa representing 34 phyla. The character matrix was analyzed by using NMDS. (*B*) Morphospace of the Chordata, which includes 26 vertebrata taxa, 2 urochordate taxa, and a single cephalochordate taxon grouped by class. (*C*) Morphospace of the 94 arthropod taxa included in the study grouped by subphylum.

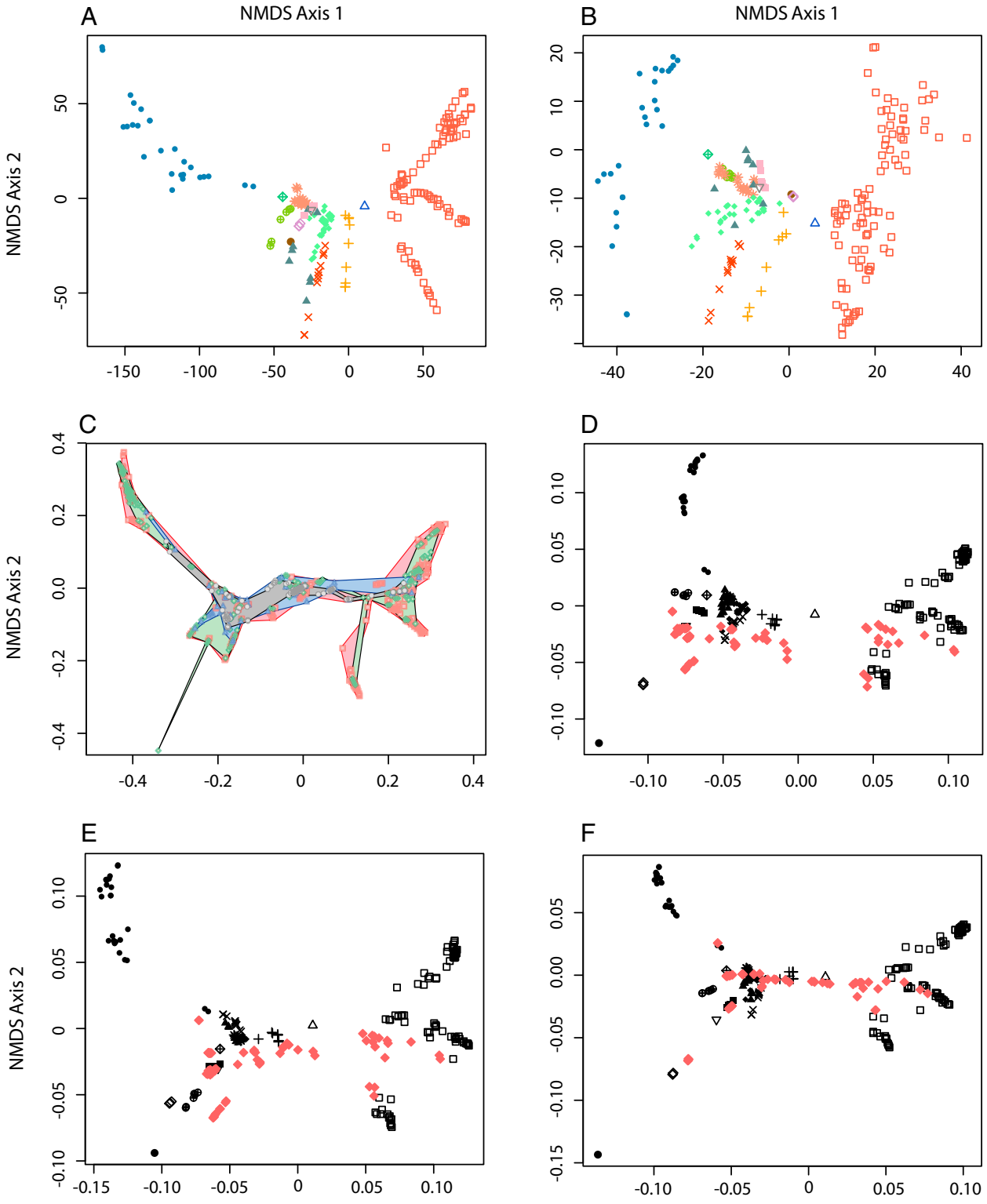


Fig. 2. Exploration of the impact of ordination method, coding strategy, character fossilization potential, inclusion of fossil taxa, and controlling for missing data. (*A* and *B*) The effect on the morphospace with different treatment of nonapplicable data. (*A*) Morphospace constructed by coding nonapplicable characters as absent. (*B*) Morphospace constructed by coding nonapplicable characters as missing. (*C*) Ordination based on a Q-mode analysis of the character matrix, considering character variance based on their taxonomic distribution; characters are color-coded according to the taxonomic rank at which they exhibit greatest variance (black, greater than phylum; blue, phylum; green, subphylum to class; red, lower than class). (*D*–*F*) Incorporation of fossil taxa (fossils represented as red diamonds; extant taxa black following symbol scheme in Fig. 1). (*D*) Addition of fossil taxa with unknown character states treated as missing data. (*E*) Impact of the loss of nonpreservable characters on morphospace structure built by using 1,000 characters that were identified as preservable based on known fossils examples or theoretical preservabilty of the structures being characterized. (*F*) Addition of fossil taxa onto the morphospace in which missing data has been modeled based on their phylogenetic position. All morphospaces were constructed similarly to Fig. 1 by using NMDS and Gower's similarity metric, with the exception of *A* and *B*, which used Manhattan distance.

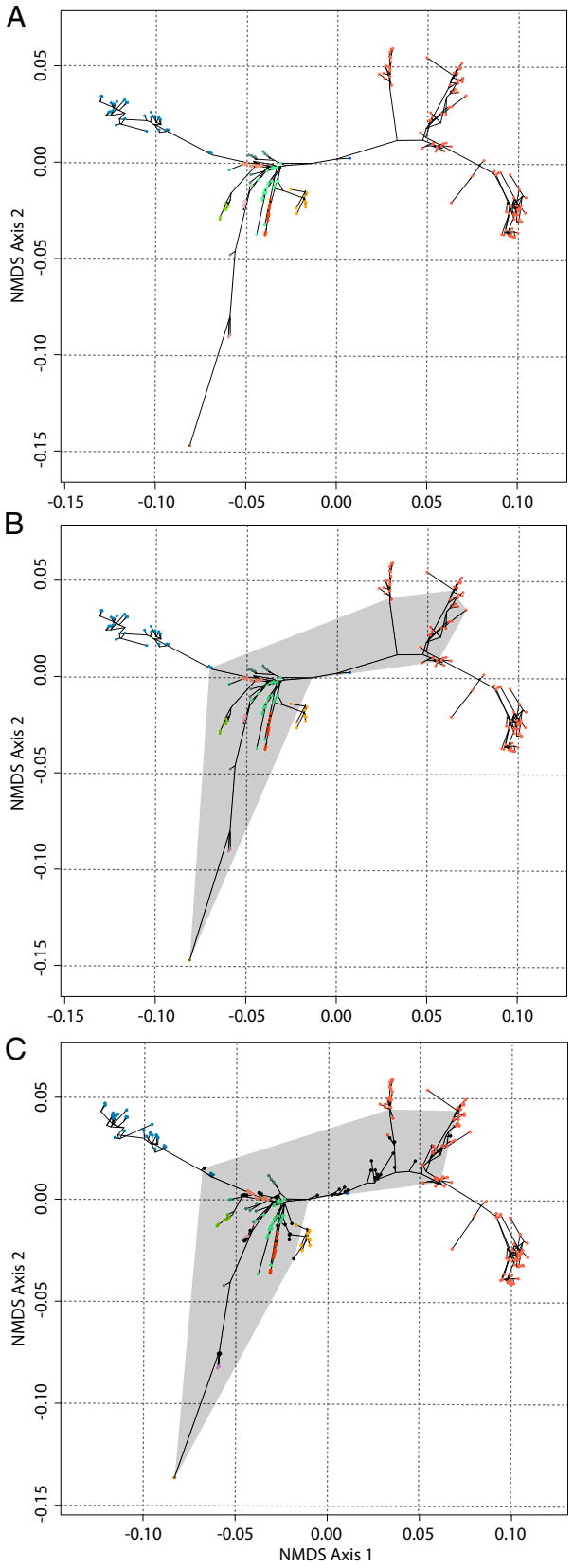


Fig. 3. Phylomorphospace and circumscription of Cambrian vs. recent animal disparity. (*A*) Phylomorphospace derived by using a consensus phylogenetic tree for the included extant taxa (*SI Appendix*, Fig. S2) and character states inferred for all of the internal nodes and tips by using stochastic character state mapping. (*B*) Convex hull (gray) circumscribing clades established before the end of the Cambrian based on fossil and molecular clock data (29). (*C*) Phylomorphospace incorporating the earliest (Cambrian) representatives of animal orders, as identified in ref. 26, with a convex hull (gray) circumscribing clades established before the end of the Cambrian based on fossil ages and molecular clock data (29); the fossil taxa were included in the ordination, but this has little qualitative impact on the distances exhibited by extant taxa. Cambrian organisms are represented by black nodes, whereas the color scheme for extant taxa follows Fig. 1*A*.

S4 and S5). We completed the coding for the fossils through stochastic character mapping (27, 28), a probabilistic approach that accommodates the uncertainty in ancestral and tip states, based on current hypotheses of their phylogenetic position (*SI Appendix*, Fig. S5 and Dataset S2). On average, we were able to code 45.4% of the characters based on fossil material, and, of the remaining 54.6% of modeled characters, 98.8% were modeled as absent or nonapplicable. The majority of the traits that were inferred to be present from the character modeling are traits that are shared by all bilaterians. Fossil taxa were then included in the ordination among their extant relatives (Fig. 2*F*); comparison with the ordination of extant taxa alone (Fig. 1*A*) shows that their inclusion does not have a significant or qualitative impact on the universe of empirical morphospace defined by the living taxa ($P = 0.001$). We also reconstructed ancestral character states for all of the internal nodes in the tree, which represent hypothetical ancestors (Dataset S2), by using the same method of stochastic character state mapping (27, 28), and plotted the phylogeny into morphospace (Fig. 3*A*). We also subdivided the nodes (internal and terminal) into Cambrian and post-Cambrian origination (based on the fossil records of lineages and divergence time estimates for some ancestral nodes from ref. 29) to assess the scale of pre-Ordovician vs. post-Cambrian innovation (Fig. 3*B*).

Superimposition of the tree topology (Fig. 3*A*) reveals that living clades do not deviate significantly from the paths that animal phylogeny is inferred to have coursed through morphospace, and extinct taxa (fossils and internal nodes) plot intermediate of their living relatives. These results support the view that the apparent distinctiveness of phylum-level crown groups and, more generally, the "clumpiness" of animal disparity, are consequences of the extinction of phylogenetic intermediates. By implication, the aspect of morphological disparity recognized by the Linnean ranks (e.g., ref. 9) is largely an artifact of later Phanerozoic extinction, not of late Neoproterozoic–Cambrian innovation. For example, the addition of the Cambrian stem-arthropods, *Anomalocaris*, *Aysheaia*, and *Opabinia*, does expand the envelope of morphospace occupied by the arthropods, but does so by bridging the gap to onychophorans (Fig. 3*C*). Hence, the distinctiveness of panarthropod phyla has increased over time with the extinction of these now-"stem" arthropods. There is no evidence, however, that the overall envelope of metazoan morphospace occupation has diminished significantly as a consequence of extinction since the Cambrian, nor that maximal disparity was achieved early in animal evolutionary history (compare Fig. 3*B* vs. Fig. 3*C*). Quite to the contrary, the inclusion of these Cambrian arthropods only expanded the region and density of morphospace occupied by arthropods, and only then in diminishing the distance between arthropods and their nearest living relatives, the onychophorans. With the exclusion of Cambrian vertebrates, the inclusion of Cambrian taxa does not in itself increase the envelope of net metazoan morphospace, which remains defined by living clades. The envelope of metazoan disparity expanded post-Cambrian, and numerous reversals are represented by crossing evolutionary pathways in Fig. 3. Reversals, and the obvious overlap in morphospace occupation by the majority of phyla, reflect the role of convergence and constraint in metazoan diversification (30). Evidently, there is no general trend in the

tempo of clade disparity: the majority of phylum-level clades exhibit maximal disparity achieved by the Cambrian (Fig. 3, gray), compatible with previous studies at low taxonomic rank (18, 31, 32), whereas others exhibit a progressive exploration of morphospace: principally arthropods and chordates (corroborating refs. 33–35), but also annelids, echinoderms, and mollusks. Thus, the envelope of disparity explored by Kingdom Metazoa has increased through geological time.

Relationship Between Disparity and Complexity, Body Size, and Diversity. Having codified metazoan disparity, we next attempted to understand the relationship between morphology and other primary biologic metrics. To achieve this, taxonomic rank had to be normalized to the phylum level. The morphologic position of phyla was determined by using two methods: (*i*) including the crown ancestor of each phylum in the preexisting morphospace and (*ii*) independently analyzing the modeled characters for the crown ancestor of each phylum (*SI Appendix*, Fig. S6). The first method is influenced by differences in diversity or disparity between phyla, whereas the second disregards those differences between phyla in an effort to control for potential sampling biases. As expected, the two morphospaces differ in the uniqueness of arthropods and chordates, which alter the strength of correlation between morphology and some other datasets. However, these two methods maintain the structure of the morphospace (Mantel test, $R = 0.946$ $P = 0.001$; *SI Appendix*, Fig. S6). Thus, differences in sampling are not the controlling factor in correlations between morphology and other datasets. In all of the following tests of correlation, we undertook parallel analyses by using both methods, the results of which were similar; we present only those results based on the independent ordination of the character sets inferred for the phylum crown-ancestors. The distances between the morphological position of the phyla are then considered as a qualitative measure of the overall similarity of the phyla and can be used as a guide to compare morphology to other primary biological metrics.

The concept of disparity has been linked to, and sometimes even conflated with, the concept of organismal complexity (36). To explore their relationship, we compiled a new dataset of metazoan cell type diversity, the only widely accepted proxy for morphological complexity (e.g., ref. 11). We questioned whether the two phenomena were correlated by using a nonlinear (Spearman) Mantel test to compare pairwise distances derived from the morphology and complexity datasets (Table 1). The possibility that the two datasets are uncorrelated can be rejected at a high level of significance (Table 1). This relationship can be rationalized because only simple body plans are possible with few cell types, compatible with the view that expansion in cell type diversity has underpinned the expansion of metazoan disparity (37).

Morphologic distances between phyla were also compared with compilations of minimum, maximum, and range in body size within each phylum (38). Body size is correlated with many ecological and evolutionary traits and has increased by 16 orders of magnitude during the history of life (39). Morphologic distance correlates significantly with maximum body size and range of body size, but not with minimum body size (Table 1). This correlation reflects the greater physical demands and adaptive solutions to body form required by larger body size. It also suggests that there is a threshold in body size below which broad phenotypic disparity may not be possible, perhaps linked to the greater diversity of cell types that characterize organisms that achieve large body size (40, 41).

The relationship between diversity and disparity has been an area of intense study in deciphering the meaning of the two metrics (42) as well as the use of higher-level diversity as a proxy for disparity (9). Many metrics have been used to calculate disparity from constructed morphospaces, but the average squared distance between taxa within morphospace shows the greatest stability with smaller sample sizes (43). Differences in modern species-level diversity within phyla (22) correlate to morphologic distances between phyla (Table 2), and the number of species within a phylum correlates strongly to the disparity contained within (Fig. 4*A* and Table 2). This indicates that, at a higher taxonomic level, these metrics are closely related. However, a comparison of phylum-level diversity and disparity through time (based on origination data from ref. 29) indicates that there is no correlation between these two aspects of variance. Indeed, the relationship is static through time (Fig. 4*B*), indicating that the number of phyla provides a poor measure of metazoan disparity. However, our results cannot reject equivalence between our measure of disparity and diversity measured by counts of Linnean ranks below the phylum level. In sum, Linnean rank taxonomic measures of disparity have overestimated the scale of early metazoan disparity and therefore the phenomenon to be explained by intrinsic and extrinsic causal factors.

Table 1. Statistical comparison of morphology against other biologic features

Feature	Mantel *r*	*P* value	No. of phyla
No. of cell types	0.3387	0.01	29
Minimum body size	0.09537	0.248	28
Maximum body size	0.4875	0.004	28
Range in body size	0.4862	0.007	28
Species level diversity	0.3678	0.007	32
Genome length	0.278	0.028	22
Protein (superfamily)	0.2512	0.141	12
Protein (family)	0.1701	0.218	12
Protein (architecture)	−0.288	0.886	12
microRNA	0.39	0.005	24

The positions of phyla were calculated in two manners. First, the modeled character suites for the ancestral nodes of the individual phyla were projected onto the morphospace constructed with Ax's 212 operational taxa. This method includes the disparity contained within phyla to some degree and therefore could be potentially biased by differential sampling. Second, the modeled character suites for the ancestral nodes were independently ordinated by using NMDS. This method treats each phylum equally and disregards the synapomorphies contained within each phylum such that the structure of the data are not controlled by differential sampling. The results were the same, so only the later is presented. Correlations between matrices were analyzed by using Mantel tests (Spearman), which compare the rank-order distance between phyla within the two matrices. The number of phyla included in the test varies with the availability of data (lists of phyla included in the different comparisons are presented in *SI Appendix*, Table S3). Multivariate datasets (miRNA and proteins) were analyzed in a similar manner to the morphological dataset (NMDS). Body size data are from McClain and Boyer (38). Genome length data were accessed from the animal genome size database

Testing Hypotheses of Causality. Distilling the phenomenon of animal disparity is one thing; establishing its causality is another. Explanations encompass intrinsic causes, including expansions in genome size (44), the diversification of protein domains and domain architectures (45), the origin of a "developmental toolkit" of transcription factors and cell signaling molecules (46), and the evolutionary assembly of gene regulatory networks (GRNs) (5), although it has also been argued that the exploration of metazoan morphospace is largely a time-dependent random walk (36). To test among these hypotheses, we compiled datasets of phylum origination dates (29), protein domains and their architectures (47), average genome sizes (48), and microRNAs (49) to serve in proxy for the diversity of GRNs. The multivariate protein and microRNA data were also analyzed by using NMDS. Our tests are limited to correlation with the use of a Mantel test. The position of a phylum within morphospace was taken as the position of its crown ancestor (*SI Appendix*, Fig. S6).

Table 2. Statistical comparison of morphology against diversity and origination

Diversity/origination	Spearman's ρ	*P* value	No. of phyla
Species level diversity	0.698	<0.001	32
Species level diversity (excl. 0s)	0.865	<0.001	16
Origination	0.041	0.905	11
Origination (excl. 0s)	0.047	0.903	9

The positions of phyla were calculated as described in Table 1. The number of phyla included in the test varies with the availability of data (lists of phyla included in the different comparisons are presented in *SI Appendix*, Table S3). Morphological disparity was calculated as the average squared distance between taxon within a phylum. Diversity was compiled by Chapman (22), and origination data are from Erwin et al. (29)

To explore whether the observed patterns in phylum-level morphologic differences could be the result of accumulated random evolutionary processes over time (36), we compared origination time to within-phylum disparity (Fig. 4*C*). In the absence of morphological constraint or stabilizing selection, clades that originated earlier would have more time to accumulate random changes and therefore to accumulate a higher level of disparity. No significant trend is seen between disparity and origination time (Table 2), indicating that, even though random processes may be contributing to disparity, there must be other, dominant factors that cause the vast difference in morphologic expansion between phyla regardless of origination time.

To test among the intrinsic causes of disparity (*SI Appendix*, Tables S3 and S4), we first compared differences in genome size (c-value) to the distances between phyla within morphospace. Genome size has been linked to several biological features such as organismal complexity, metabolic rates, and developmental rate (50), and we can reject the hypothesis that genome size and morphological complexity are uncorrelated at a high level of significance (Table 1). Genome size is not correlated to the number of genes (50), suggesting that the link between genome size and morphology might instead be effected largely through gene regulation, rather than the number of genes per se. Although the expansions of some protein-domain superfamilies have been correlated with the evolution of complexity (51), our data show no overall correlation between disparity and the repertoire of protein domains or architectures (Table 1 and *SI Appendix*, Fig. S7). In contrast, the number of miRNA families correlates significantly with morphological disparity (Table 1 and *SI Appendix*, Fig. S8), lending correlative support to the hypothesis that expansions in gene regulatory complexity underlie the evolution of metazoan morphological complexity (5, 7, 52).

Discussion

Our mapping of metazoan morphospace has shown that, even though some animal phyla demonstrate maximal initial disparity, others, notably chordates, arthropods, annelids, and mollusks, have progressively expanded on the limits of phylum and kingdom level morphospace post-Cambrian. This contradicts the generally held view of maximal initial disparity for the animal kingdom, which has been based largely on two assumptions: (*i*) that Linnean ranks serve as an effective proxy for organismal disparity and (*ii*) that patterns of morphospace occupation captured by continuous or discrete character observations at low taxonomic rank reflect the same phenomenon at phylum and kingdom levels. These are assumptions we refute. Not all studies have returned a pattern of maximal initial disparity. Indeed, disparity analyses based on continuous (53) and discrete characters (18) have identified instances in which maximal disparity is achieved in the middle period or late in the evolutionary history of lineages. Nevertheless, recent compilations have shown that, more often than not, maximal disparity is achieved early in the evolutionary history clades (10, 18). However, it does not follow logically that, at the comparatively low taxonomic level at which these studies have been conducted, a pattern of maximal initial disparity will scale to a self-similar pattern at the highest taxonomic levels. Nevertheless, disparity at low and high taxonomic levels is linked hierarchically such that, regardless of whether its zenith is achieved early or late, increasing disparity at low taxonomic

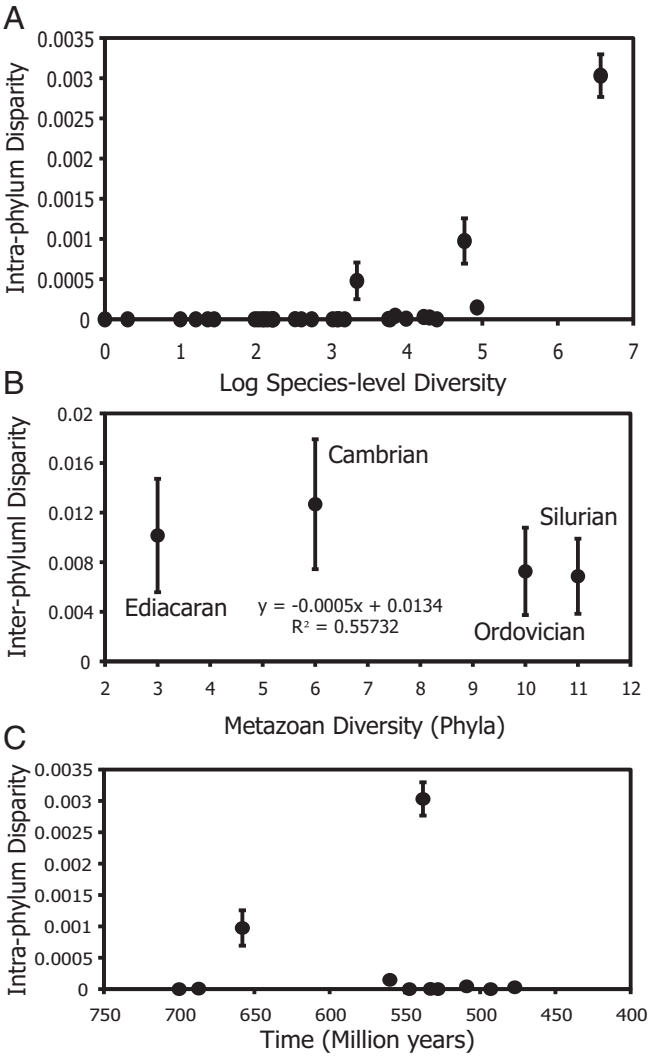


Fig. 4. Comparisons between taxonomic diversity and morphologic disparity. Disparity is calculated as the average squared distance between taxa within the morphospace. Intraphylum disparity is calculated by using the distance between Ax's operational taxa within the morphospace for each phylum, and interphylum disparity is calculated by using the distance between phyla (*SI Appendix*, Fig. S5). (*A*) Intraphyla disparity compared with estimates of modern diversity. If there was a single operational taxon within a phylum, the disparity is given a value of 0. (*B*) Comparison between metazoan diversity and interphylum disparity for 4 time intervals for the 11 phyla with well-resolved origination dates (*SI Appendix*, Table S5). For each time bin, the number of phyla that were present during the beginning of the interval were tallied (i.e., metazoan diversity) so, for example, 3 of the 11 phyla originated before the beginning of the Ediacaran, whereas an additional 3 phylum-level crown groups had appeared by the beginning of the Cambrian. Interphylum disparity was then calculated for the phyla present during that interval. (*C*) Intraphyla disparity compared with the age of the phylum. Origination data are from Erwin et al. (29) and diversity data are from Chapman (22). Error bars are calculated as the SE of 1,000 bootstrap replicates.

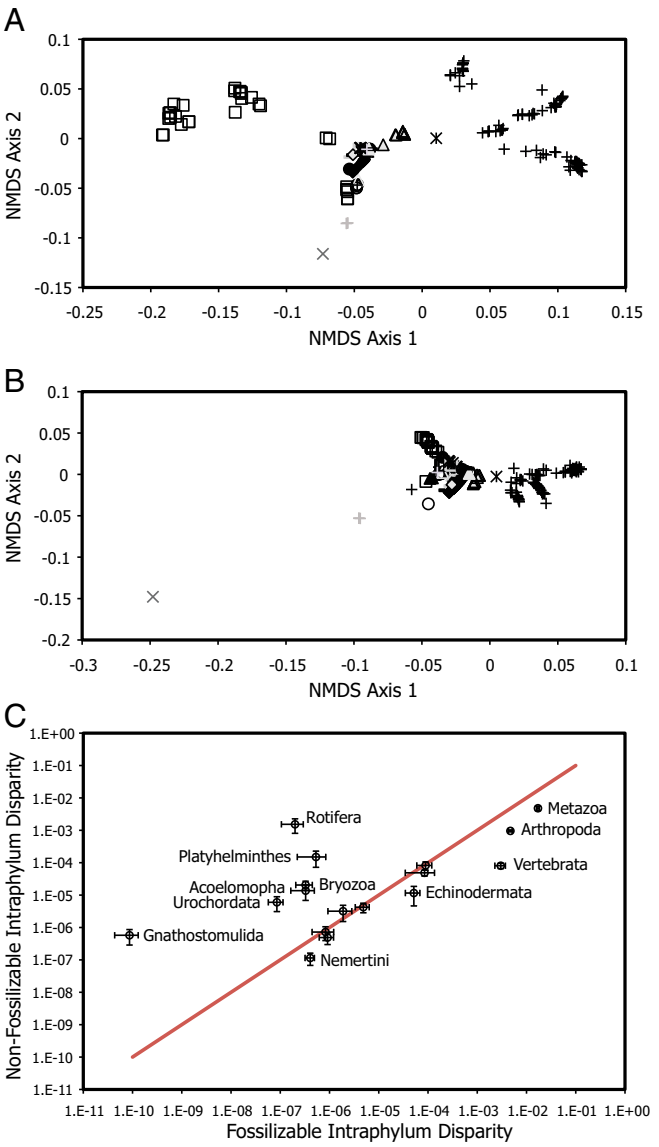


Fig. 5. Distribution of taxa within morphospace based on fossilizable and nonfossilizable characters. (*A*) Morphospace based on 912 fossilizable characters. (*B*) Morphospace based on 878 characters unlikely to be observable in fossil organisms. (*C*) Intrametazoan and intraphylum disparity based on ordinations of only fossilizable characters and only nonfossilizable characters. Phyla above red line have higher disparity within nonpreserved soft-tissue characters, whereas those below have higher disparity within preserved gross anatomical or skeletal characters.

levels must scale to gross disparity increasing progressively at the highest taxonomic levels.

The clumpiness of metazoan morphospace occupation has been alternately argued to be a consequence of (*i*) gene regulatory and/or other constraints such as integration or modularity, (*ii*) extinction, and (*iii*) incomplete exploration of morphospace. Our analyses demonstrate that, even though many (but not all) modern clades occupy discrete regions of morphospace, the inclusion of Cambrian taxa indicates that phylogenetic intermediates of living clades occupy concomitantly intermediate regions of morphospace. Combined with our inferences of the course of metazoan phylogeny through morphospace, these results indicate that the morphological discreteness of modern clades is largely a consequence of the extinction of phylogenetic intermediates. Nevertheless, the results of our analyses including fossils and phylogenetic ancestors indicate that the majority of the empirical morphospace circumscribed by our dataset has not been explored in metazoan evolutionary history. This may be because insufficient time has elapsed for all of these theoretical phenotypes to have been realized during metazoan phylogeny. However, there is a high level of phylogenetic linkage among many of the phenotypic characters in the dataset (Dataset S1): approximately half of the characters in the dataset are contingent directly on just 20 characters, and almost all of the characters are contingent ultimately on just a few characters (epithelia; ontogenesis; somatic differentiation). This reflects a high level of phylogenetically rooted developmental constraint underpinning the distribution of characters and their possible combinations. Thus, it is likely that the majority of the theoretical morphologies represented by unoccupied volumes of morphospace are unrealizable because the implied character combinations are not possible (e.g., ref. 54).

Evidently, the heterogeneity in the phylogenetic linkage of characters reflects the fact that some have contributed more than others to the realization of metazoan bodyplans. These characters can be identified based simply on their contingent linkage, and also through ordination of the phenotypic characters based on their distribution among species rather than through ordination of the species based on their distribution among characters (Fig. 2*C*). This reveals that the large distance of morphospace that separates nonmetazoans from metazoans is influenced strongly by those characters with a heavy contingent burden that evolved early within the metazoan stem-lineage. Crown metazoan characters, such as a defined head, stomochord, protocoel, ecdysis, segmentation, and features of nephridial cells, all load strongly along NMDS axis one, indicating that these characters are important in the establishment of metazoan body plans.

Undoubtedly, it would be useful to explore the impact of including a greater number of fossil taxa in future analyses of

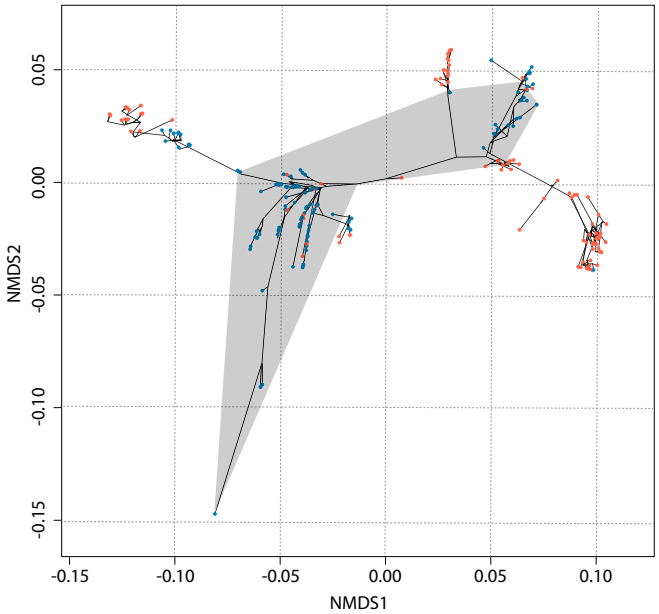


Fig. 6. The role of transitioning to terrestrial environments in the post-Cambrian exploration of morphospace. The phylogenetic pathway within morphospace as well as the area occupied during the Cambrian is shown as in Fig. 3. Organisms that are exclusively aquatic (marine, brackish, freshwater, or parasitic) are colored in blue, whereas taxonomic groups that have some terrestrial representation are colored in red. Terrestrial groups are included within arthropods, chordates, onychophorans, tardigrades, gastropods, rotifers, annelids, nematodes, nemerteans, and Platyhelminthes.

evolutionary disparity that build upon the dataset we present. However, fossil species will always be limited in the amount and class of data they contribute to disparity analyses. Reducing the dataset to features that can be preserved in the fossil record fails to encompass the richness of characters that diagnose the bodyplans of high-rank taxa, and the ordination of such a dataset results in a perspective on metazoan diversification that is reduced dramatically, demonstrating that many of the key superphylum characteristics defining axis one are lost taphonomically. Even though the structure of the morphospace is retained by using preserved or nonpreserved subsets of the characters, there are major differences between the resulting ordinations (Fig. 5). This is a consequence of an increase in apparent disparity within highly skeletonized phyla (Arthropoda, Vertebrata, and Echinodermata) at the expense of those phyla that are extensively soft-bodied (e.g., Rotifera, Platyhelminthes, and Gnathostomulida). The absence of soft-tissue characteristics exaggerates the disparity within the independently derived skeletal features, masking the differences between the majority of phyla, which are based largely on soft-tissue characters, and diminishing the scale of the morphological radiation within metazoans as a whole. Therefore, it is likely that, because the anatomical features that underlie the deep connections within metazoans are not preserved, trends in metazoan disparity cannot be addressed with fossil taxa alone.

The results of this analysis also suggest that the perception of metazoan diversification represented in the fossil record serves to exaggerate the increase in disparity associated with the origin of fossilizable characters. Thus, the "Cambrian Explosion" phenomenon may more represent an explosion in fossilizable characters, and therefore fossils (55), rather than the dramatic increase in phenotypic disparity it has long been interpreted to represent. In demonstrating a dramatic post-Cambrian expansion in the maximum variance in phenotypic disparity, our results indicate that the scale of the Cambrian Explosion has been grossly overestimated. This effectively mirrors the extensive pre-Cambrian evolutionary history of metazoans estimated by molecular clock methodology (56), in providing more time for the accrual of metazoan disparity achieved before the end of the Cambrian and thereby making the challenge of identifying causality far more tractable.

Our tests of hypotheses of causality cannot reject a role for random variation in effecting metazoan phenotypic disparity, but we find no correlative support for the role of expansions in the protein-coding repertoire of the genome. However, our results lend correlative support to the hypothesis that expansions in gene regulatory complexity underlie the evolution of metazoan morphological complexity. This does not preclude a role for extrinsic mechanisms such as the expansion of ecospace (57). Indeed, the phyla that deviate most significantly from the thesis of maximal initial disparity (arthropods, mollusks, annelids, chordates) did so in large part within the hitherto underexplored terrestrial environment (Fig. 6). In so doing, these lineages were released from the physical constraints of an aqueous marine environment to explore new realms of morphospace. Although the causal factors underlying metazoan organismal disparity have been considered nonuniformitarian (7, 8), our analysis shows that the capacity for novelty in metazoan evolution has not disappeared since the Cambrian. Of course, this may be because the intrinsic processes generating the underlying genetic and developmental variation are the same.

Our results also suggest that debate on whether early animal evolution has been underpinned by uniformitarian or non-uniformitarian processes has been misplaced. Animal evolutionary history does not appear to have been characterized by a uniform rate and scale of change but rather by a high frequency of small changes and low frequency of changes of large magnitude within the context of intrinsic genetic and developmental variation and extrinsic environmental change. Such patterns are readily open to modeling in the same manner as nucleotide and amino acid substitution frequencies. Future research in this direction will inform understanding of the nature of phenotypic evolution, its relation to molecular evolution, underpinning the development of phylogenetic methods. However, it will also provide for a more precise characterization of the tempo of metazoan diversification and the processes that underpinned the establishment of animal bodyplans.

ACKNOWLEDGMENTS. We thank Neil Shubin (Chicago), Peter Wagner (Lincoln, NE) and two anonymous referees for their patience in assisting in the development of the manuscript. This work was funded through grants from the Natural Environment Research Council (to P.C.J.D.), the National Science Foundation (to B.D. and K.J.P.), the NASA National Astrobiology Institute (to K.J.P.), and a Benjamin Meaker Visiting Professorship from the Institute of Advanced Studies, University of Bristol (to B.D.).

1. Gould SJ (1990) *Wonderful Life: The Burgess Shale and the Nature of History* (Hutchinson Radius, London).
2. Briggs DEG, Fortey RA, Wills MA (1992) Morphological disparity in the cambrian. *Science* 256:1670–1673.
3. Foote M (1997) The evolution of morphological diversity. *Annu Rev Ecol Syst* 28:129–152.
4. Hall BK (1996) *Baupläne*, phylotypic stages, and constraint. Why there are so few types of animals. *Evol Biol* 29:215–261.
5. Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. *Science* 311:796–800.
6. Davidson EH, Erwin DH (2010) Evolutionary innovation and stability in animal gene networks. *J Exp Zoolog B Mol Dev Evol* 314:182–186.
7. Erwin DH, Davidson EH (2009) The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* 10:141–148.
8. Erwin DH (2011) Evolutionary uniformitarianism. *Dev Biol* 357:27–34.
9. Erwin DH, Valentine JW, Sepkoski JJ (1987) A comparative-study of diversification events–The early Paleozoic versus the Mesozoic. *Evolution* 41:1177–1186.
10. Erwin DH (2007) Disparity: Morphologic pattern and developmental context. *Palaeontology* 50:57–73.
11. Villier L, Eble GJ (2004) Assessing the robustness of disparity estimates: The impact of morphometric scheme, temporal scale, and taxonomic level in spatangoid echinoids. *Paleobiology* 30:652–665.
12. Foth C, Brusatte SL, Butler RJ (2012) Do different disparity proxies converge on a common signal? Insights from the cranial morphometrics and evolutionary history of Pterosauria (Diapsida: Archosauria). *J Evol Biol* 25:904–915.
13. Hopkins MJ (2017) How well does a part represent the whole? A comparison of cranidial shape evolution with exoskeletal character evolution in the trilobite family Pterocephaliidae. *Palaeontology* 60:309–318.
14. Hetherington AJ, et al. (2015) Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology* 58:393–399.
15. Huttegger SM, Mitteroecker P (2011) Invariance and meaningfulness in phenotype spaces. *Evol Biol* 38:335–351.
16. Mitteroecker P, Huttegger SM (2009) The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors. *Biol Theory* 4:54–67.
17. Gerber S (2016) The geometry of morphospaces: Lessons from the classic Raup shell coiling model. *Biol Rev Camb Philos Soc* 92:1142–1155.
18. Hughes M, Gerber S, Wills MA (2013) Clades reach highest morphological disparity early in their evolution. *Proc Natl Acad Sci USA* 110:13875–13879.
19. Ax P (1996) *Multicellular Animals: A New Approach to the Phylogenetic Order in Nature* (Springer, Berlin), Vol 1, p 225.
20. Ax P (2000) *Multicellular Animals: The Phylogenetic System of the Metazoa* (Springer, Berlin), Vol 2, p 369.
21. Ax P (2003) *Multicellular Animals: Order in Nature–System Made by Man* (Springer, Berlin), Vol 3, p 317.
22. Chapman A (2005) *Numbers of Living Species in Australia and the World* (Australian Biological Resources Study, Canberra, Australia), p 61.
23. Gower J (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27:857–871.
24. Deline B (2009) The effects of rarity and abundance distributions on measurements of local morphological disparity. *Paleobiology* 35:175–189.
25. Deline B, Ausich W (2011) Testing the plateau: A reexamination of disparity and morphologic constraints in early Paleozoic crinoids. *Paleobiology* 37:214–236.
26. Tweedt S (2013) First appearences of major metazoan clades in the fossil record. *The Cambrian Explosion: The Construction of Animal Biodiversity* (Roberts, Greenwood Village, CO), pp 343–354.
27. Huelsenbeck JP, Nielsen R, Bollback JP (2003) Stochastic mapping of morphological characters. *Syst Biol* 52:131–158.
28. Revell LJ (2012) Phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 3:217–223.

29. Erwin DH, et al. (2011) The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. *Science* 334:1091–1097.
30. Conway Morris S (2003) *Life's Solution: Inevitable Humans in a Lonely Universe* (Cambridge Univ Press, Cambridge, UK), p 650.
31. Wills M (1998) Cambrian and recent disparity: The picture from priapulids. *Paleobiology* 24:177–199.
32. Foote M (1999) Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. *Paleobiology* 25:1–116.
33. Stockmeyer Lofgren A, Plotnick RE, Wagner PJ (2003) Morphological diversity of Carboniferous arthorpods and insights on disparity patterns of the Phanerozoic. *Palaeobiology* 29:350–369.
34. Wagner PJ, Ruta M, Coates MI (2006) Evolutionary patterns in early tetrapods. II. Differing constraints on available character space among clades. *Proc Biol Sci* 273:2113–2118.
35. Sallan LC, Friedman M (2012) Heads or tails: Staged diversification in vertebrate evolutionary radiations. *Proc Biol Sci* 279:2025–2032.
36. McShea DW, Brandon RN (2010) *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems* (Univ Chicago Press, Chicago), p 170.
37. Valentine JW, Collins AG, Meyer CP (1994) Morphological complexity increase in metazoans. *Paleobiology* 20:131–142.
38. McClain CR, Boyer AG (2009) Biodiversity and body size are linked across metazoans. *Proc Biol Sci* 276:2209–2215.
39. Payne JL, et al. (2009) Two-phase increase in the maximum size of life over 3.5 billion years reflects biological innovation and environmental opportunity. *Proc Natl Acad Sci USA* 106:24–27.
40. Bonner JT (1965) *Size and Cycle: An Essay on the Structure of Biology* (Princeton Univ Press, Princeton, NJ), p 260.
41. Bonner JT (1988) *The Evolution of Complexity* (Princeton Univ Press, Princeton, NJ).
42. Foote M (1993) Contributions of individual taxa to overall morphological disparity. *Paleobiology* 19:403–419.
43. Ciampaglio CN, Kemp M, McShea DW (2001) Detecting changes in morphospace occupation patterns in the fossil record: Characterization and analysis of measures of disparity. *Paleobiology* 27:695–715.
44. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
45. Tordai H, Nagy A, Farkas K, Bányai L, Patthy L (2005) Modules, multidomain proteins and organismic complexity. *FEBS J* 272:5064–5078.
46. Carroll SB (1995) Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–485.
47. de Lima Morais DA, et al. (2011) SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic Acids Res* 39:D427–D434.
48. Gregory TR (2012) Animal Genome Size Database. www.genomesize.com.
49. Tarver JE, et al. (2013) miRNAs: Small genes with big potential in metazoan phylogenetics. *Mol Biol Evol* 30:2369–2382.
50. Gregory TR (2005) Genome size evolution in animals. *The Evolution of the Genome*, ed Gregory TR (Elsevier, San Diego), pp 3–87.
51. Vogel C, Chothia C (2006) Protein family expansions and biological complexity. *PLOS Comput Biol* 2:e48.
52. Peterson KJ, Dietrich MR, McPeek MA (2009) MicroRNAs and metazoan macroevolution: Insights into canalization, complexity, and the Cambrian explosion. *BioEssays* 31:736–747.
53. Foote M (1992) Paleozoic record of morphologica diversity in blastozoan echinoderms. *Proc Natl Acad Sci USA* 89:7325–7329.
54. Foote M (1994) Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. *Palaeobiology* 20:320–344.
55. Runnegar B (1982) The Cambrian explosion–animals or fossils. *J Geol Soc Aust* 29: 395–411.
56. dos Reis M, et al. (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr Biol* 25:2939–2950.
57. Erwin DH (1994) Early introduction of major morphological innovations. *Acta Palaeontol Pol* 38:281–294.